Livestock market data for modeling disease spread among US cattle

Ian T. Carroll*[1,2] and Shweta Bansal[1,2]

[1] *Department of Biology, Georgetown University, Washington, DC, USA*
[2] *Fogarty International Center, National Institutes of Health, Bethesda, MD, USA*

April 4, 2016

Abstract

Transportation of livestock carries the risk of spreading foreign animal diseases, leading to costly public and private sector expenditures on disease containment and eradication. Livestock movement tracing systems in Europe, Australia and Japan have allowed epidemiologists to model the risks engendered by transportation of live animals and prepare responses designed to protect the livestock industry. Within the US, data on livestock movement is not sufficient for direct parameterization of models for disease spread, but network models that assimilate limited data provide a path forward in model development to inform preparedness for disease outbreaks in the US. Here, we develop a novel data stream, the information publicly reported by US livestock markets on the origin of cattle consigned at live auctions, and demonstrate the potential for estimating a national-scale network model of cattle movement. By aggregating auction reports generated weekly at markets in several states, including some archived reports spanning several years, we obtain a market-oriented sample of edges from the dynamic cattle transportation network in the US. We first propose a sampling framework that allows inference about shipments originating from operations not explicitly sampled and consigned at non-reporting livestock markets in the US, and we report key predictors that are influential in extrapolating beyond our opportunistic sample. As a demonstration of the utility gained from the data and fitted parameters, we model the critical role of market biosecurity procedures in the context of a spatially homogeneous but temporally dynamic representation of cattle movements following an introduction of a foreign animal disease. We conclude that auction market data fills critical gaps in our ability to model intrastate cattle movement for infectious disease dynamics, particularly with an ability to addresses the capacity of markets to amplify or control a livestock disease outbreak.

*Corresponding author: *itc2@georgetown.edu*

Author Summary

We have automated the collection of previously unavailable cattle movement data, allowing us to aggregate details on the origins of cattle sold at live-auction markets in the US. Using our novel dataset, we demonstrate potential to infer a complete dynamic transportation network that would drive disease transmission in models of potential US livestock epidemics.

Introduction

Livestock operations within the United States (US) must be vigilant against transboundary animal diseases, including the critical threat to agriculture posed by a re-introduction of foot-and-mouth disease (FMD) [1]. Livestock movements were critical to the experience of the United Kingdom (UK) in 2001, where FMD was also non-endemic and where measures to control its outbreak included a temporary, nation-wide ban on the transportation of any livestock [2]. During the so-called "silent spread" period before the movement ban, nine out of the 12 geographic clusters of the epidemic had already been exposed to the disease through movement of infected animals between livestock operations [3]. Eventual costs to the agricultural sector reached £3 billion, and 5% of the UK's 11 million cattle were culled to control the disease [2, 4]. In the US, potential impacts of livestock movement on an FMD outbreak among the 90 million cattle maintained for beef and dairy production [5] are not easily estimated, but one study with disease spread constrained to California's 5 million beef and dairy cattle predicts economic losses in the tens of billions of dollars with hundreds of thousands of culls [6].

The spread of disease between livestock operations through movement of infected animals is an inherent risk of modern systems of meat and dairy production in the US [7]. Many factors influence the progression of cattle from widely dispersed calving operations to relatively few packing facilities, including long term trends in productivity and consumer demand, a complex decadal-scale cycle in cattle inventories, and biologically driven seasonal cycles [8,9]. Variation in the volume of livestock in transit can be epidemiologically significant: for example, one factor amplifying the 2001 epidemic of FMD in the UK was the timing of the outbreak, which coincided with a seasonal peak in the number of sheep moving through livestock markets [3]. The number of animals moved between operations doubly affects the consequences of an outbreak, once through direct increases in opportunities for disease transmission and again for any additional time involved in tracing the movements of infected animals. Delays in the discovery of infected premises, on the order of hours rather than days for a disease as virulent as FMD, are certain to impact epidemic size and economic costs [6].

To understand the implications of livestock movement on the potential for disease spread in the US, it is necessary to consider the options available to livestock producers who buy and sell cattle. Auction markets (also referred to as stockyards and sale barns) are the predominate method of sale for producers selling cattle and calves in the US, although privately negotiated sales, packer-owned feedlots and other alternatives are common [10]. Auction market sales involve the movement of cattle to a stockyard, where they are grouped into lots, sold and transported to

various buyers; alternative marketing methods avoid stockyards and vertically coordinate production chains [11], reducing the potential for contact between susceptible and infectious livestock. Producers selling through auction markets additionally influence the potential spread of a disease through choice in marketing areas, which tend to be large and overlapping [12, 13]. Despite extensive economic theory on when, where and how producers ought to market cattle, no data on real movements to or from auction markets have been collected to inform epidemiological analysis in the US.

FMD preparedness warrants close attention to auction markets, which played a central role in early, rapid expansion of the 2001 FMD epidemic in the UK [14] and possibly in several other countries with recent outbreaks of non-endemic FMD [15]. Epidemiological contact rates with auction markets in parts of Colorado and Kansas [16], California [17], and Texas [18] have been estimated through industry participant surveys. The most recent national survey indicates that over half of beef producers sold non-breeding stock at auction markets in 2006-07 (60.7% of producers for steers, 58.3% for cows [19]). Dairy contributes fewer US cattle shipments, but cows removed from dairy operations are also predominantly marketed at a stockyard [20]. While such surveys indicate the volume of cattle moved through particular marketing areas, they do not resolve movements between specific buyer or seller locations and individual stockyards, the kind of data increasingly in demand for emerging, epidemiological models for the spread of livestock disease [21].

Network models provide a layer of abstraction between raw data on individual contact rates, here cattle movements, and disease models [22]. Their use is primarily motivated by heterogeneity in the number of disease transmitting contacts attributed to infectious nodes (i.e. individuals or sub-populations), a pattern that emerged strongly during the initial spread of FMD in the UK's 2001 epidemic [14]. Transportation networks for several European livestock industries [23–28] are known because animal tracking systems are mandated by the European Parliament [29]. Networks inferred from these data have lead to several results on surveillance and control strategies; for example, (1) identification of "sentinel" livestock premises projected to become infected early during an outbreak in Italy [30], (2) validation of risk reduction from the standstill rules implemented in the UK after 2001 [31], and (3) evaluation of targeted movement bans that selectively eliminate high risk contacts [28]. Network models for the UK cattle production system have additionally provided a foundation for livestock transportation strategies that promise efficient control of *endemic* diseases [32].

Livestock movement networks in the US have previously been inferred from interstate certificates of veterinary inspection (ICVIs), which provide the only regularly-collected record of cattle movement in the US [33–35]. However, ICVIs are only reported for interstate movement when only 20% of shipments onto US beef operations originate over 100 miles away [36]. Additionally, ICVIs are not required under federal rules for cattle moved interstate to approved livestock marketing facilities (including auction markets, which can play an outsized role in disease spread) when their next movement is intrastate [37].

Here, we introduce a novel stream of data on livestock movement through auction markets in the US, summarize the record of actual movements collected to date, and implement an approach to fitting a dynamic network model for livestock

movement to this dataset. The data source is auction markets themselves, some of which publicly share the location of origin for representative lots of cattle sold on a particular day as advertisement for their business. Estimating a network model requires two levels of extrapolation from these data. First, we demonstrate that a sampling model allowing for strong overdispersion extrapolates locations of origin among representative lots to locations for all cattle received at auction. Second, we relate the patterns of movement to spatial covariates from the agricultural census, demonstrating potential to infer unobserved movements of cattle to auction markets that do not issue public market reports. We conclude with a model implementing the spread of disease on a livestock movement network; the model demonstrates how market biosecurity practices impact epidemic prevention in the aftermath of a foreign animal disease introduction. Animal tracing rules promulgated by the US Department of Agriculture (USDA) [37] forestall detailed movement tracking for predictive disease modeling, so we anticipate that our data mining approach will continue to be the only avenue for recording actual, intrastate and auction market movements of cattle in the US.

Methods

As an integral part of the livestock system, auction markets are distributed throughout the US (Box 1A) [38]. USDA market reporters attend sales (i.e. the live auction of multiple lots of cattle occurring on a particular day) for many auction markets in order to collect volume and price information that is distributed through the USDA Agricultural Marketing Service's Market News (MN). To provide additional information for advertising purposes, market reports may be published on websites maintained by individual auction markets. These independent market reports include one or more of: the total number of cattle sold which is referred to as "receipts", price ranges for different types of cattle, and the winning bid for particular "representative lots" of cattle. Descriptions of representative lots include the source of cattle, the type and number of cattle and the price obtained. The total number of cattle within representative lots rarely approaches receipts in any market report, so both receipts and representative lots are relevant to estimation of cattle movement networks.

We define the epidemiologically-relevant contact network as a bipartite network between county nodes (containing multiple farms) and market nodes, connected by edges representing movement of cattle between a county (or the farms in it) and a given auction market (Box 1B). Our data result from egocentric sampling on the market nodes. Receipts correspond to the rate of flow (i.e. node strength) through a given market node; and representative lots represent a sample of movement (i.e. edge weights) from origin to destination nodes (Box 1C).

In the first section below, we describe the aggregation of data compiled from market reports across market-owned websites. The following section explains methods for estimation and validation of a network model, built on the newly collected data and for the purpose of driving national-scale cattle disease simulations. These methods lay the groundwork for a network model serving to extrapolate from our data to a complete network of market-directed cattle movement in the US. Data on

receipts and representative lots collected for this study are freely available online (Supplementary Data).

Data collection

We automated the aggregation of receipts and representative lots across livestock auctions held at various auction markets in the US (Box 1A). Overlapping and incomplete directories of auction markets are maintained by multiple regulatory agencies and one business association. We compiled four such directories to identify auction markets [38] with potential to publish representative lots. The Livestock Marketing Association [39] uniquely provides websites of auction markets, so we manually searched the 322 websites included in the LMA directory for market reports. We wrote software in Python to download and parse information from websites that regularly (usually weekly) publish market reports including representative lots and which permit crawling by the Robots Exclusion Protocol. For each market report, the software cross-references the date against the market's corresponding series of reports available through the MN search [40] to relate our records to a MN report on the same sale, when available.

For each representative lot in a market report, the software parses the consignor's (i.e. seller's) location, cattle type (e.g. steer or heifer), number (or "head") of cattle, their average weight, and the purchase price (either per head or per hundred weight). It also extracts total receipts from any market report that includes it as well as receipts from the corresponding MN report if available. Where necessary, market reports are converted to plain text format from images with Tesseract OCR [41] or from PDFs with pdftotext [42]. A single animal per lot is assumed whenever the number of cattle cannot be parsed. We tuned the parser for each website until two researchers confirmed 100% accuracy of representative lots in at least two current market reports and a haphazard sample of archived reports, which varied widely in availability. Websites continue to be checked twice weekly for new reports, and parsers failing to obtain weekly data, returning data of the wrong type (i.e. string or numeric), or returning a head total from representative lots that exceeds receipts are manually checked and promptly corrected. The present analyses incorporate market reports obtained from June 2014 through Feb 28, 2016, including archived reports on sales dating from the first week of 2012.

The location of origin for a representative lot is typically given as a city or other populated place, with or without a state, and can be ambiguous. On the assumption that livestock operations are proximal to, but rarely within, the given city, our software georeferences each city to a US county, parish or equivalent (hereafter "county"). This loss of geographic precision is primarily a trade-off for accuracy, but the county scale also aligns with administrative boundaries relevant to disease management. The software matches locations, with common abbreviations expanded as needed, against names of populated places in North America using the GeoNames web-service [43] in order to identify the encompassing county. The county closest to the reporting market's county, as determined by the great-circle distance (GCD) between county centroids [44], is recorded as the true location of origin. Our analyses include any market report with at least one representative lot whose origin could be georeferenced to a county by this procedure.

In order to extrapolate from collected market reports to unobserved movements of cattle, we incorporate cattle and calf inventory and sales from county tables in the USDA National Agricultural Statistics Service's Census of Agriculture [5] as independent variables. After eliminating one of any pair of variables with strong correlation apparent on logged or untransformed axes, the following independent variables for each county remain: inventory of beef cows (IB), inventory of milk cows (IM), inventory of cattle on feed (IF), and sales by number of cattle over 500lbs (SC). Data for counties with few operations is often redacted, but we filled in most of the missing data by 1) summing over the product of the state average value per operation with the number of operations in the county for each operation size category, and then 2) subtracting non-redacted components from aggregated fields (e.g. inferring milk cow inventories as the difference between total cow and beef cow inventories). We created binary variables (e.g. $nzIB$ for non-zero inventory of beef cows) to distinguish zero from non-zero values within each numeric variable and transformed the latter by taking z-scores of the logarithm of non-zero values (e.g. $lnIB$). A final group of variables for each county and market pair describes whether the county and market are in the same state (S), whether the market is in the county (C), the z-score of the non-zero distance between centroids of the county or origin and the market's county (D), and the z-score of the logarithm of this distance (lnD).

Inferring unobserved movements

We define our network model as a joint probability distribution over the edges between pairs of nodes, representing counties and auction markets. Each edge corresponds to the number of cattle from a given county (of c total counties) that are sold at a given auction market (of m total markets). A realization of the network model is the set of non-negative integers $n_{i,j,t}$ representing movement of cattle to market i from county j at each time period t (i.e. the total number of cattle on *one* of the left-hand edges in Box 1C). For each realization, the corresponding value for receipts consigned at market i at time t is $r_{i,t} = \sum_{k=1}^{c} n_{i,k,t}$. We employ a joint probability distribution over network edges of the form

$$\prod_{t} \prod_{i=1}^{m} P_i \left(n_{i,1,t}, n_{i,2,t}, \ldots, n_{i,c,t} | r_{i,t}\right) P_{i,t}(r_{i,t}), \tag{1}$$

which embeds several simplifying assumptions. The foremost assumption is independence, both between markets and over time, for all edges representing cattle movements. Another implied assumption is that time-dependence only factors into the distribution on receipts, $P_{i,t}(\cdot)$, or the total number of cattle sold at market i during period t. In other words, the number of cattle from each county has a constant probability distribution conditioned on receipts in this model. Choosing a model of this form facilitates estimation of two distinct factors, $P_i(\cdot|\cdot)$ and $P_{i,t}(\cdot)$. We employ the first as a likelihood over representative lots and the second as a likelihood over receipts. We fit parameters of the likelihood models, described next, by sampling posterior parameter vectors with Hamiltonian Monte Carlo implemented in Stan [45] with weakly-informative Gaussian or half-Cauchy priors (Supplementary Methods).

The multivariate distribution on the number of cattle from each county, $P_i(\cdot|\cdot)$, has the sample space of a multinomial distribution, but the sampling method dictates a different likelihood. The variable represents a sum over the number of cattle in one or more representative lots from each county, and not, as in a multinomial distribution, a sum over independent trials with equal probability. The Pólya (a.k.a Dirichlet-multinomial) distribution, parameterized by non-negative $\boldsymbol{\alpha}_{i,\cdot}$, flexibly admits the overdispersion resulting from this "bursty" sampling method [46]. We initially explored the suitability of using the Pólya family in our likelihood function by directly estimating $\boldsymbol{\alpha}_{i,\cdot}$ for each auction market independently, i.e. including only the ith market's representative lots in the likelihood. Subsequently, we estimate posterior distributions over parameter vector $\boldsymbol{\beta}_x$ affecting all markets and counties simultaneously through linear combination with independent variables $\boldsymbol{x}_{i,j}$. In this case, the first factor in the network model (1) has definition

$$P_i\left(\boldsymbol{n}_{i,\cdot,t}|r\right) = \mathcal{P}\left(\boldsymbol{n}_{i,\cdot,t}|\alpha_{i,1} = e^{\boldsymbol{x}_{i,1}\boldsymbol{\beta}_x}, \alpha_{i,2} = e^{\boldsymbol{x}_{i,2}\boldsymbol{\beta}_x}, \ldots, \alpha_{i,c} = e^{\boldsymbol{x}_{i,c}\boldsymbol{\beta}_x}\right), \tag{2}$$

where $\mathcal{P}$ is the Pólya distribution's probability mass function. Note that for model fitting, we take $n_{i,j,t}$ to be a number of cattle observed in representative lots, for which the sum across counties is just the total head in representative lots and not the market's receipts for that auction. Only when using (1) predictively do we take $n_{i,j,t}$ to equal all receipts from the jth county. The independent variables in $\boldsymbol{x}_{i,j}$ have fixed effects $D_{i,j}$, $lnD_{i,j}$, $S_{i,j}$, $C_{i,j}$, $nzIB_j$, $nzIM_j$, $nzIF_j$, $nzSC_j$, $lnIB_j$, $lnIM_j$, $lnIF_j$, $lnSC_j$, and a random effect of market, $M_i \sim \mathcal{N}(0, \sigma_x)$. Posterior predictive modeling, using MCMC samples of $\boldsymbol{\beta}_x$, is possible for any market for which $\boldsymbol{x}_{i,j}$ are available, which admits non-reporting markets in any censused county.

The network model (1) gains temporal dynamics by allowing the distribution on receipts, $P_{i,t}(\cdot)$, to change over time. The volume of cattle traded at specific auction markets is driven by social and economic variables at several spatial and temporal scales, so we do not presently attempt to model receipts at auctions for which we did not obtain a market report. Instead, we model receipts for auction markets in relation to the number of cattle from all representative lots in a given market report, on the assumption that this total follows the trends in receipts. For receipts $r_{i,t}$ at market i having representative lots totaling $s_{i,t}$ head during time period t, we assume a negative-binomial distribution on the unobserved head, $r_{i,t} - s_{i,t}$. Parameterizing its probability mass function by the mean μ and a parameter ϕ such that the variance is $\mu + \mu^2/\phi$, we set the second factor in the network model (1) to

$$P_{i,t}(r_{i,t}) = \mathcal{NB}\left(r_{i,t}|\mu = e^{\boldsymbol{y}_{i,t}\boldsymbol{\beta}_y}, \phi\right). \tag{3}$$

Posterior samples for $\boldsymbol{\beta}_y$ and ϕ are obtained using the subset of market reports for which we obtained receipts. The independent variables in $\boldsymbol{y}_{i,t}$ represent fixed effects for year (Y_t) and week of year (W_t) along with $s_{i,t}$, $lnIB_i$, $nzIM_i$, $nzIF_i$, $lnIM_i$, $lnIF_i$, $lnSC_i$, and a random effect of market with variance σ_y^2.

Validation

Markets contributing to the data stream are self-selecting, prompting the need for validation against randomly sampled data where possible. In addition, the number

of cattle transported to a given auction from each county is a high-dimensional response variable with potential to violate the assumptions of our multivariate regression model. We validate the likelihood function assumed in our Bayesian inference by checking self-consistency: we test the goodness-of-fit between posterior predictive distributions and observed data. Specifically, we estimate posterior predictive distributions for two summaries of the data, the number of counties represented at a given market averaged across market reports and the average distance cattle are moved to a given market, and compare to empirical distributions using the one-sample KS test. To seek evidence against biases in the underlying data itself, we compare our movement data to previous estimates of livestock transportation. Freely available data on livestock movements in the US is only available for year 2001 health-certificate derived data on interstate flows [7], but these data have recently been validated against ICVIs from year 2009 [33]. We aggregated all representative lots from a full year of market reports by the markets' state, and correlated the number of cattle by state of origin with year 2001 interstate flows.

Results

We report on sales from Jan 2, 2012 to Feb 28, 2016, with a sample size of 13 markets in 2012 increasing to 63 markets by 2016 (Fig. 1A). Cattle from representative lots in 5171 reports from 63 markets located in 60 counties in 17 states are represented in this analysis (Box 1A). The total number of cattle included among 347 thousand representative lots is 3.1 million. In the three sections which follow, we summarize overall trends observed in the data aggregated across market reports, show the predictability of the county of origin for cattle sold at different markets, and relate the number of unobserved cattle sold at a given auction to both static and time-varying covariates.

Static and seasonal patterns

The average number of cattle movements reported in representative sales for a live auction increases to a peak of 1000-1500 head in late fall and decreases to a few hundred during summer months (Fig. 1B). This seasonal trend persists for the duration of our study period. For a given week, livestock markets report receiving cattle from 11 (SD 1.5) counties on average, and this lower bound on network "in-degree" also shows seasonal variation with increases of a lesser magnitude in late autumn (Fig. 1C). Seasonal peaks for in-degree roughly follow the variation in cattle volumes, but grow less pronounced as the number of markets included in the dataset grows. Likewise, the number of counties represented during low seasons fails to track the volume of cattle before the most recent year, but the larger samples achieved since late 2015 show in-degree variation that is well correlated with variation in cattle volumes. The average distance between county of origin and auction market for cattle in representative lots is well under 100 miles in the vast majority of weeks (Fig. 1D), which s consistent with producer surveys [36].

Interstate movements are scarce among representative lots, but those observed are well correlated with pre-existing state in-shipment data. Aggregating across all markets, the proportion of sales that originate within-state shows no temporal

Table 1. Correlations between interstate representative lots and inshipment data from Shields & Mathews [7]. For states with at least one reporting market, the table shows the Pearson correlation (ρ) for the number of cattle from n origin states, along with the number of cattle in the representative lots that moved between states and its proportion (p) of the total.

Dest.	p	Head	n	ρ
IA	0.99	20	34	-0.04
AZ	0.99	25	15	-0.10
ID	0.66	7174	43	0.25
KS*	0.95	5600	38	0.37
CO*	0.85	10915	23	0.48
MO	0.75	11544	17	0.49
MN*	0.96	4082	37	0.54
CA	1.00	19	43	0.66
WY*	0.86	10072	28	0.72
OK	0.68	6381	18	0.73
NV	0.97	166	28	0.77
ND	0.98	4300	3	0.78
SD	0.74	110644	39	0.78
TX*	0.98	287	33	0.79
MT*	0.92	16628	16	0.90
NM	0.83	4403	17	0.99

* State import rules, as presented on each state's agricultural department website, except certain livestock markets from ICVI requirements.

trend in deviations from an average of 0.84 (SD 0.067). Oklahoma and South Dakota auction markets typically sold the most out-of-state cattle, and were the only states with a market that on-average advertised more out-of-state cattle than in-state cattle (Fig. 1E). Nonetheless, among the 16 states (Tennessee excluded) with year 2015 representative lots that could be compared to year 2001 in-shipments [7], the Pearson correlation coefficients for cattle by origin state are greater than 0.48 in all but two reliable estimates (Table 1). Representative lots marketed in Iowa, Arizona, and California were insufficient for meaningful comparison, and the lowest remaining coefficient (0.25 for Idaho) is driven by one strong connection to Nye County in Nevada. South Dakota has the largest number of interstate representative lots that we observed and has a strong correlation of 0.79. Montana and Colorado, while having similarly large sample sizes, exhibit opposite extremes with the first highly correlated at 0.90 and the other relatively weakly at 0.48.

Fitting to representative lots

A first test of the network model is whether any parameterization of the Pólya family can fit the high-dimensional data on the number of cattle by county for each market, leaving aside the regression against independent variables. The posterior mean of $\boldsymbol{\alpha}_{i,\cdot}$ (the Pólya parameters for market i) indicates the fitted strength of as-

sociation between each market and different counties (Fig. 2A), but doesn't directly correspond to the observed average number of cattle for that edge because of variation in the total number of cattle between market reports. A better visualization of the goodness-of-fit is the posterior predictive distribution on the number of counties represented as a function of this total (Fig. 2B), which falls below what would be expected under a multinomial sampling model whenever cattle from the same county are clustered in lots. These fitted curves capture the trend and variability in the county-distribution of representative lots for all but a handful of markets with smaller sample sizes (Supplementary Figures).

Markets naturally exhibit variation in the weight of edges connecting to each county, quantified by the posterior mean of the normalized $\boldsymbol{\alpha}_{i,\cdot}$, as well as variation in the degree to which representative lots approach independent samples across counties, measured by the sum of $\alpha_{i,j}$ over j (Fig. 2C). Mapping the edges visualizes the sharp decline in weight with greater distance, anticipating a strong effect of distance when the $\alpha_{i,j}$ are related to independent variables. Variation among markets in the total of $\boldsymbol{\alpha}_{i,\cdot}$ indicates differences in overdispersion relative to a multinomial distribution, with a smaller total indicating more "bursty" arrivals of cattle from each county. Only a handful of markets have relatively large totals, indicating that most counties with farms consigning cattle to these few markets are represented at every sale. In the majority of markets, counties are more irregularly represented at a given sale, but show up with larger numbers of cattle when present. Another visually apparent distinction between markets is on the distribution of weights among edges pointing to a given market: the market in Nevada, for example, has a long tail of infrequently represented counties when compared to the market in Arizona.

In the second round of fitting to representative lots, using likelihood (2) to incorporate all representative lots conditioned on the total per market report, we predict edge weight and node strength with independent variables related to geography and agricultural production. Samples from the posterior distribution on $\boldsymbol{\beta}_x$ are negative for both linear and logarithmic functions of distance (D 95% CI -1.90 to -1.86, lnD 95% CI -1.313 to -1.306) as anticipated. The county in which a market resides is typically well represented at auctions there (95% CI 10.6 to 10.7), and administrative boundaries also amplify expected movements to in-state markets (95% CI 0.76 to 0.78). Agricultural census data included in the model have different effects based on the class of cattle. The presence and abundance of cows in a county increased its edge weights ($nzIM$ 95% CI 0.11 to 0.13, IM 95% CI 0.066 to 0.071, $nzIB$ 95% CI 3.1 to 3.3, IB 95% CI 1.00 to 1.02), with beef cows having the larger effect. In contrast, the presence and abundance of cattle that are further along the chain of production (e.g. feeder cattle) reduced edge weights ($nzIF$ 95% CI -0.20 to -0.18, IF 95% CI -0.084 to -0.075, $nzSC$ 95% CI -4.01 to -3.88, SC 95% CI -0.35 to -0.34). We capture unexplained, between-market variation in the total of $\boldsymbol{\alpha}_{i,\cdot}$ with a random intercept (σ_x 95% CI 1.07 to 1.15); however, the posterior mean coefficients indicate some pressure to deviate from the assumption of a normal distribution (Fig. 3A).

A comparison between observed data and posterior predictive distributions indicates both strengths and weaknesses in the regression model. The in-degree distribution across markets (averaged over market reports) does not deviate from the distribution generated by the assumptions of the fitted likelihood model (KS statis-

tic 0.064, $P > 0.1$). In other words, given a total number of cattle in representative lots consigned at a given auction, the model groups them into the correct number of counties on average (Fig. 3C). However, the distribution across markets on the average distance cattle move to market deviates from the prediction (KS statistic 0.41, $P < 0.001$). There is greater variability among markets in the average distance that cattle are transported to market (Fig. 3D) than the regression model achieves, indicating the need for an interaction between distance-related covariates and individual markets.

Fitting to receipts

The likelihood on receipts captures the dynamic nature of the movement network by positively responding to variation in the total number of cattle in representative lots and including non-zero coefficients for different weeks and years (e.g. Fig. 4A). Two distinct processes represented by these coefficients drive temporal variation in receipts within a market. On average, auction markets report representative lots in proportion to the receipts for a given sale, posting more representative lots when more cattle are sold (95% CI 0.47 to 0.49). This proportion, however, is not constant over time as additional increases in un-observed receipts are associated with certain weeks (e.g. 95% CI 0.36 to 0.52 in late October) and decreases occur in others (e.g. 95% CI -0.59 to -0.38 for late July). The time-varying fixed effects combine with the negative binomial overdispersion parameter (ϕ 95% CI 4.3 to 4.4) to produce realistic levels of within market variation in the proportion of receipts appearing in representative lots (Fig. 4B); the remaining independent variables, which also happen to be static, account for between market variation. The occasional underestimate of receipts during peak cattle trading season could likely be resolved by relating ϕ to some of these same independent variables.

Agricultural census data for the county in which an auction market is located statistically explains part of the variation between markets, and the remainder is attributed to a random effect of market. Non-zero inventories of milk cows (95% CI -3.5 to -2.7) and cattle on feed (95% CI -2.5 to -1.4) both reduce the expected number of un-observed receipts, and the magnitude of cattle inventories or sales have further negative effects ($lnIM$ 95% CI -0.45 to -0.35, $lnIB$ 95% CI -0.79 to -0.55, $lnIF$ 95% CI -0.83 to -0.65, $lnSC$ 95% CI -0.40 to -0.086). Each market has a random intercept with standard deviation σ_y (95% CI 0.68 to 0.73) that is large compared to the standard deviation among markets for the combined effect of the census variables (95% CI 0.010 to 0.12), implying that much of the variation in the unobserved receipts (Fig. 4B) is "explained" by unknown sources of variation between markets. In addition, for a handful of markets, the static census data for that county drives the predictive distribution for un-observed receipts well below the true values, suggesting that some critical process is not represented in the covariates (Supplementary Figures).

Application: Modeling Disease Spread

To illustrate the utility of our network model, we study infection propagation from county to county, thwarted by market biosecurity measures. For this, we develop

a simple infection spread model driven by variability in the number of cattle arriving at market from a county. In conjunction with movement bans for designated "control areas" [35] that serve as the primary strategy for containment of an FMD outbreak in the US [1], biosecurity procedures at auction markets are a critical mechanism for preventing movement of potentially infectious animals between livestock operations. The premise is that biosecurity practices at auction markets can provide a firewall, preventing infection from reaching new operations by intercepting cattle that may be carrying disease from an operation with infected animals. Our model tracks the number of counties (we assume that movement bans are implemented one whole county at a time) that must be designated as control areas during eradication of an FMD-like outbreak. All exports from a control area are restricted, but movements that occur before the control is established are responsible for growth in the number of counties affected. The way an area becomes eligible for control is through a "dangerous contact", which we define as receiving cattle from any farm in an area that previously had a dangerous contact or was infected. We assume here that all movements occur through a livestock market, and that contact between animals at a livestock market does not engender additional dangerous contacts between operations. In summary, a county becomes a control area if it receives a shipment of cattle from any county that is also destined to become a control area, but has not become one yet.

We model disease spread through a branching process [47 section 15.2] that extends n generations if spread is halted by the time any one county is at most n dangerous contacts removed from the county at the source of the outbreak. The model is completely specified by choosing a distribution on the number of distinct counties that could have a dangerous contact with a single county. We denote this random variable by Y and estimate its distribution using the above likelihood models, with parameters sampled from estimated posterior distributions given the market data, and additional assumptions (Supplementary Methods). Biosecurity practices at auction markets are incorporated into the distribution on Y as a zero-inflation parameter. With probability τ, a market quarantines cattle arriving from a farm within a county that will become a control area, resulting in no additional control areas becoming established due to consignments at that sale. The remaining parameters in our model may be explained through their effect on the expectation $\langle Y \rangle = \lambda(1-\tau)\nu\langle N \rangle$: λ is an average number of marketings, ν is a splitting parameter for the allocation of cattle from the origin county among destination counties, and $\langle N \rangle$ is the expected number of cattle per origin county per sale. Then, the expected value of the total number of counties, X_n, that must become control areas in a branching process stopped after n generations, is

$$\langle X_n \rangle = \frac{1 - \langle Y \rangle^{n+1}}{1 - \langle Y \rangle}. \tag{4}$$

Additional details in the Supplementary Methods justify this result under the strong assumption that the distribution on Y is the same for any county which becomes a control area, which is only reasonable for small n.

Variation in the factors contributing to $\langle Y \rangle$ drive the effort needed to eradicate the disease; in particular, they determine where the critical threshold lies that separates a finite branching process from a potentially infinite one. The key contribution

from a network model fitted to market report data is estimates of the number of cattle consigned at a given sale from farms within a single county. The estimate must be extrapolated from a combination of representative lots and receipts, which the fitted network model (1) rightfully predicts to be greater than direct observations (Fig. 5A). Temporal variation in the number of cattle marketed per county, which is slight but still apparent in both the observed and predicted distributions, is primarily due to variation in total cattle volumes (Fig. 1B). These volumes fluctuate more than the number of counties represented at auctions (Fig. 1C), which would otherwise have a counter-effect on the number marketed per county at different times of year. For fixed values of λ and ν, weekly variation over the course of the year in cattle sales strongly affects the threshold value of τ separating the two qualitatively distinct outcomes of the branching process (Fig. 5B). During more active cattle trading seasons, biosecurity must be considerably stronger to cause convergence of X_n as $n \to \infty$ or, equivalently, to bring the spreading process to a stop without further intervention. Averaging over the seasonal variation in cattle trade, for the purpose of considering introduction of an FMD-like disease at a randomly selected time of year, the effect of market biosecurity strongly influences the magnitude of any resulting eradication effort. The value of $\langle X_n \rangle$ is the minimum number of control areas to establish if back-tracing to the origin from the county of detection involves a sequence of n dangerous contacts. Thus, interpreting n as a measure of the delay between disease introduction and a regional movement restriction that halts further dangerous contacts, we see that around the critical value of τ, doubling the delay nearly doubles the number of control areas needed (Fig. 5C).

Discussion

The epidemiological contact network is a fundamental component of models for the spread of a disease, and market reports publicized by livestock auction markets can contribute urgently needed data to support inference of such networks within the US livestock system. In this report, we describe an ongoing process to opportunistically sample cattle movements originating from counties with beef or dairy operations and consigned at auction markets distributed across the US. Our study complements previous efforts to summarize transportation of cattle within the US using data derived from interstate certificates of veterinary inspection [7, 34], but is novel because it includes within-state movements and targets movements involving auction markets, which can be crucial to amplification or control of a foreign animal disease. We develop a statistical model for inference of a bipartite contact network between nodes representing either a US county (containing cattle farms) or a US cattle auction market, and demonstrate its link to epidemiological models of economically disruptive livestock diseases.

Data on cattle movements that are voluntarily reported by auction markets have not previously been aggregated and collectively analyzed, for either economic or epidemiological purposes. In contrast, data on volumes and prices from auction markets is central to the economic analysis of livestock systems, is naturally of keen interest to industry participants, and its reporting is mandated by law [48]. A few studies have examined data on movement of cattle sold through auctions [12, 49],

but those studies focus on internet auctions which do not imply movement of cattle through a physical stockyard. It is critically important to archive data released by auction markets while it is still available, and in particular before components of this novel data stream dry up. Recent declines in the reliance on auction markets for slaughter cattle have become severe enough to raise new concerns about price discovery, the backbone of an efficient beef market [48]. While such declines affect the economically important class of slaughter cattle, reliance on auction markets for cattle with epidemiological significance, those moving between farms (e.g. replacements) or to feeding operations, shows no sign of decline (Supplementary Figures).

Cattle tracing systems in the UK and Europe create opportunities for research on a fully censused network, but the situation in the US is such that relevant network data must be sparsely sampled and produces a challenging population inference problem. In network sampling theory, randomly selecting nodes without tracing their edges to sample additional nodes, is an ego-centric sampling method that allows straightforward estimation of population level node attributes, but not of edge attributes [50]. In contrast, were *edges* randomly sampled as in the ICVI records, the reverse problem applies: the sample of nodes involved is automatically biased towards those of higher degree. Sampling theory pertaining to higher-order network properties, such as clustering and assortativity, becomes exceedingly complicated. Lacking a universally applicable sampling method for network data, what's needed are computationally feasible methods for estimating the complete network in a manner that respects the likelihood of observing each network attribute with a given sampling strategy [51]. Our study lays the groundwork for extrapolating a complete network of cattle transported to auction markets from an ego-centric sample of representative lots and receipts at individual market nodes. While we might reasonably estimate average market degree (i.e. the number of connected counties) with a simple sample mean, we would not, for example, estimate the expected number of cattle moved on any edge in the complete network with a simple mean across representative lots. Instead, we propose a regression model with a flexible likelihood function to scale up from the known constraints of our sampling method. Our success at obtaining posterior probability distributions for the parameters relating cattle movement to independent variables is due to an efficient HMC sampler [45] in combination with an efficient algorithm for calculating a Pòlya likelihood for multiple samples (Supplementary Methods). Incorporation of variables beyond the agricultural census data is computationally feasible in this framework, which already promises reliable estimates.

Prior to model fitting, we confirm that our data is generally consistent with known patterns and trends in the livestock industry. For example, seasonal variation in the volume of representative lots is consistent with beef cattle production systems, where calves are produced in spring and weaned cattle or yearlings sold to pasturing or feedlot operations in the fall and subsequent spring [52]. In addition, the average distance most cattle in representative lots move to a market is under 100 miles, while considerably longer distance movements do infrequently arise, as in survey results [36]. Likewise, the majority of observed movements are destined to a within-state market, which is consistent with the dominance of intrastate movements extrapolated by Lindström et al [33] from entirely interstate data. As these

authors point out, the intrastate estimates could not be validated at the time; now, however, visual comparison between their state specific estimates [Fig. S4 in 33] and our data (Fig. 1E) indicates that market reports show even higher rates of intrastate movement. The un-matched data on cattle movements from market reports shows that roughly 80% of market directed shipments on any week originate within state. Among the remaining interstate movements, consignments from out-of-state counties correlate fair to strongly with state in-shipments summarized by Shields and Mathews [7]. Sale at livestock markets is not the only impetus for cattle transportation, but the correlation between representative lot origins and state in-shipments demonstrates its importance. Indeed, if the certificate-derived data do sample all movements (excepting slaughter animals) without bias, then the stronger correlations in Table 1 would indicate states for which cattle in-shipments are in fact movements to a livestock market. Alternatively, differences in ICVI requirements for cattle bound for livestock markets could affect the degree of correlation, but no such connection is apparent among the states with reporting markets.

Our confidence in the data quality extends also to the statistical model, which provides a strong starting point for predicting unobserved movements to non-reporting livestock movements. Opportunistic sampling unavoidably raises the specter of biased estimation on the population as a whole; any unknown associations between self-reporting market nodes and the volume of cattle they auction or the weight of edges from particular county nodes would bias network estimation. Regression of independent variables against receipts and representative lots incorporates some of these associations, eliminating a few potential sources of bias. More importantly, however, agreement between the sign of fitted coefficients and our intuition about the livestock system builds confidence in the model's predictive ability. The negative effect of distance, for example, is one clear example of the model following intuition. The difference between the positive effect of cows and negative effects of feeder cattle on the weight of a county's connections to markets was not anticipated, but is consistent with auction markets' role in moving livestock from cow-calf operations towards feeding operations. The regression model for receipts naturally associates larger receipts with auctions for which a greater number of cattle are reported in representative lots, but it also indicates a downward adjustment in the predicted difference between receipts and total cattle in representative lots for markets in counties with larger inventories of cattle. The latter effect implies that such markets include a greater proportion of their receipts among the representative lots–a potentially rewarding strategy for markets advertising in active livestock producing regions. Our current regression models do not address all potential sources of bias in network estimation; for example, the markets sampled are predominately located in central states and any distinction between beef and dairy cows in the data has not yet been explored. Ideally, as for validation of any statistical model fitted to mined data, or otherwise opportunistic samples, we would compare model predictions to newly collected and truly random samples of the transportation network. In the context of cattle movements, however, the privacy interests of the livestock industry are a significant obstacle to collecting such validation data.

The central weaknesses of the combination of data and network estimation reported here is whats left out, movements of cattle leaving livestock markets and extension of the model for receipts to non-reporting markets. The model for re-

ceipts introduced here relies in part on information (the total head in representative lots) only available for auctions accompanied by market reports. Unlike the conditional distribution on edges (2) then, the predictive distribution on receipts cannot immediately extend to auctions from non-reporting markets. Estimating receipts independently of partial data on a given auction requires a mechanistic model that takes into account seasonal and longer term dynamics driven by economic (e.g. prices of beef and feed) processes as well as reproductive cycles. Fortunately, there is a wealth of data on auction market receipts in the USDA Agricultural Marketing Service's Market News archives to aid in model development. A greater challenge lies in estimating the markets' out-going edges in the livestock transportation market. Information on the location or identity of cattle buyers is unlikely to be shared by any auction market, in order to protect their business. A more fruitful avenue for future research likely involves combining livestock market information with ICVI records, joining the data on both the location of origin (i.e. the market location) and the date of sale.

Auction markets can play a significant role in amplification of livestock disease, just as airports play a role for human diseases. The market report data and parameter inference performed here provide a unique resource for modeling the role market biosecurity can play, potentially bringing records of actual movements to bear on agricultural policy in ways not possible through existing [e.g. 53, 54] kernel-based models of disease spread [55]. Our study provides, for the first time, information on cattle movements that do not cross a state boundary. Although transportation of infectious cattle within a state would not immediately spark a regional epidemic, cattle movements at this scale could rapidly spread disease beyond the 10km control radius suggested in response to FMD detection within the US [1]. Our simple model of infection propagation stymied by market-based biosecurity measures highlights the tradeoffs involved in mitigating infection in the intensely-connected livestock industry of the US. In particular, we show that the efficiency of biosecurity measures at a market must be balanced against the ramifications of any delay in implementing movement bans.

Acknowledgments

The authors want to acknowledge several undergraduate or graduate assistants who contributed to software development: Daniel Anderson, Adam Graves, Ching-Hao Hu, and Xinyang Jiang. We thank Nancy Robinson at the Livestock Marketing Association for sharing a member directory, and Centennial Livestock Auction of Fort Collins, CO for answering questions on cattle market practices. We also thank Jason E. Lombard (USDA-APHIS-VS) for several helpful conversations about livestock marketing practices. Funding for contributions by ITC and SB was provided through DHS Contract #HSHQDC-12-C-0014, with additional support from the RAPIDD Program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

A Supplementary Data

Description of data release.

B Supplementary Methods

Stan model code and elaboration of branching-process model.

C Supplementary Figures

Individual market results as in Figure 2 and 4, and additional data on marketing arrangements.

References

[1] US Department of Agriculture. Foot-and-Mouth Disease Response Plan: The Red Book. USDA:APHIS:VS, FAD PReP; 2014. Available from: https://www.aphis.usda.gov/animal_health/emergency_management/downloads/fmd_responseplan.pdf.

[2] Anderson I. Foot and Mouth Disease 2001: Lessons to be Learned Inquiry Report; 2002. Available from: http://webarchive.nationalarchives.gov.uk/20100807034701/http://archive.cabinetoffice.gov.uk/fmd/fmd_report/report/index.htm.

[3] Gibbens JC, Sharpe CE, Wilesmith JW, Mansley LM, Michalopoulou E, Ryan JBM, et al. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. The Veterinary Record. 2001;149(24):729–743.

[4] Thompson D, Muriel P, Russell D, Osborne P, Bromley A, Rowland M, et al. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. Revue scientifique et technique (International Office of Epizootics). 2002;21(3):675–687.

[5] US Department of Agriculture. 2012 Census of Agriculture; 2012. Available from: ftp://ftp.nass.usda.gov/quickstats/qs.census2012.txt.gz.

[6] Carpenter TE, O'Brien JM, Hagerman AD, McCarl BA. Epidemic and economic impacts of delayed detection of foot-and-mouth disease: a case study of a simulated outbreak in California. Journal of Veterinary Diagnostic Investigation. 2011;23(1):26–33.

[7] Shields DA, Mathews KH. Interstate Livestock Movements. Electronic Outlook Report from the Economic Research Service. 2003;LDP-M-108-. Available from: http://www.ers.usda.gov/media/312234/ldpm10801_1_.pdf.

[8] Mathews KH, Hahn WF, Nelson KE, Duewer LA, Gustafson RA. U.S. Beef Industry: Cattle Cycles, Price Spreads, and Packer Concentration. Market and Trade Economics Division, Economic Research Service, US Department of Agriculture Technical Bulletin. 1999;1847.

[9] Crespi JM, Xia T, Jones R. Market Power and the Cattle Cycle. American Journal of Agricultural Economics. 2010;92(3):685–697. Available from: http://ajae.oxfordjournals.org/cgi/doi/10.1093/ajae/aap034.

[10] RTI International. GIPSA Livestock and Meat Marketing Study. Research Triangle Park, ND; 2007. Available from: https://www.gipsa.usda.gov/psp/publication/live_meat_market.aspx.

[11] RTI International. Spot and Alternative Marketing Arrangements in the Livestock and Meat Industries. Interim Report. 2005;.

[12] Bailey D, Brorsen BW, Thomsen MR. Identifying Buyer Market Areas and the Impact of Buyer Concentration in Feeder Cattle Markets Using Mapping and Spatial Statistics. American Journal of Agricultural Economics. 1995;77(2):309–318. Available from: http://ajae.oxfordjournals.org/content/77/2/309.abstract.

[13] Hirschi M, Feuz DM. Does the Law of One Price Hold for Feeder Cattle Purchased in Various Regions of the US and Shipped to One State in the Midwest? Southern Agricultural Economics Association Annual Meeting, Orlando, FL, February 6-9, 2010. 2010;.

[14] Shirley MDF, Rushton SP. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. Epidemiology and Infection. 2005;133(06):1023. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2870335&tool=pmcentrez&rendertype=abstracthttp://www.journals.cambridge.org/abstract_S095026880500453X.

[15] McLaws M, Ribble C. Description of recent foot and mouth disease outbreaks in nonendemic areas: exploring the relationship between early detection and epidemic size. Canadian Veterinary Journal. 2007;48(10):1051–1062.

[16] McReynolds SW, Sanderson MW, Reeves A, Sinclair M, Hill AE, Salman MD. Direct and indirect contact rates among livestock operations in Colorado and Kansas. Journal of the American Veterinary Medical Association. 2014;244(9):1066–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24739117.

[17] Bates TW, Thurmond MC, Carpenter TE. Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three California counties, with reference to control of po- tential foot-and-mouth disease transmission. American Journal of Veterinary Research. 2001;62:1121–1129.

[18] Ward MP, Highfield LD, Vongseng P, Graeme Garner M. Simulation of foot-and-mouth disease spread within an integrated livestock system in Texas, USA. Preventive Veterinary Medicine. 2009;88(4):286–297. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167587708002511.

[19] US Department of Agriculture. Beef 200708 Part IV: Reference of Beef Cow-calf Management Practices in the United States, 200708. Fort Collins, CO: USDA:APHIS:VS, CEAH; 2010.

[20] US Department of Agriculture. Dairy 2007 Part II: Changes in the U.S. Dairy Cattle Industry, 19912007. Fort Collins, CO: USDA:APHIS:VS, CEAH; 2008.

[21] Dawson PM, Werkman M, Brooks-Pollock E, Tildesley MJ. Epidemic predictions in an imperfect world: modelling disease spread with partial data. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1808):20150205–20150205. Available from: http://rspb.royalsocietypublishing.org/content/282/1808/20150205?etoc.

[22] Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. Journal of The Royal Society Interface. 2007 oct;4(16):879–891. Available from: http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2007.1100.

[23] Bigras-Poulin M, Thompson RA, Chriel M, Mortensen S, Greiner M. Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread. Preventive Veterinary Medicine. 2006 sep;76(1-2):11–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16780975.

[24] Kao RR, Danon L, Green DM, Kiss IZ. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. Proceedings of the Royal Society B: Biological Sciences. 2006 aug;273(1597):1999–2007. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1635475&tool=pmcentrez&rendertype=abstract.

[25] Nöremark M, Håkansson N, Lewerin SS, Lindberg A, Jonsson A. Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. Preventive veterinary medicine. 2011;99(2-4):78–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21288583.

[26] Dutta BL, Ezanno P, Vergu E. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. Preventive Veterinary Medicine. 2014;117(1):79–94. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167587714002943.

[27] Lentz HHK, Konschake M, Teske K, Kasper M, Rother B, Carmanns R, et al. Trade communities and their spatial patterns in the German pork production network. Preventive Veterinary Medicine. 2011;98(2-3):176–181. Available from: http://dx.doi.org/10.1016/j.prevetmed.2010.10.011.

[28] Natale F, Giovannini A, Savini L, Palma D, Possenti L, Fiore G, et al. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. Preventive Veterinary Medicine. 2009;92(4):341–350.

[29] Regulation (EC) No 1760/2000 of the European Parliament and of the Council of 17 July 2000 establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products and repealing Council Regulati; 2000. Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000R1760.

[30] Bajardi P, Barrat A, Savini L, Colizza V. Optimizing surveillance for livestock disease spreading through animal movements. Journal of The Royal Society Interface. 2012;9(76):2814–25. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3479905&tool=pmcentrez&rendertype=abstract.

[31] Green DM, Kiss IZ, Kao RR. Modelling the initial spread of foot-and-mouth disease through animal movements. Proceedings of the Royal Society B: Biological Sciences. 2006;273(1602):2729–35. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1635508&tool=pmcentrez&rendertype=abstract.

[32] Gates MC, Woolhouse MEJ. Controlling infectious disease through the targeted manipulation of contact network structure. Epidemics. 2015;Available from: http://linkinghub.elsevier.com/retrieve/pii/S1755436515000304.

[33] Lindström T, Grear DA, Buhnerkempe MG, Webb CT, Miller RS, Portacci K, et al. A bayesian approach for modeling cattle movements in the United States: scaling up a partially observed network. PloS One. 2013;8(1):e53432. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3537632&tool=pmcentrez&rendertype=abstract.

[34] Buhnerkempe MG, Grear DA, Portacci K, Miller RS, Lombard JE, Webb CT. A national-scale picture of U.S. cattle movements obtained from Interstate Certificate of Veterinary Inspection data. Preventive Veterinary Medicine. 2013;112(3-4):318–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24035137.

[35] Buhnerkempe MG, Tildesley MJ, Lindström T, Grear DA, Portacci K, Miller RS, et al. The impact of movements and animal density on continental scale cattle disease outbreaks in the United States. PloS One. 2014;9(3):e91724. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3966763&tool=pmcentrez&rendertype=abstract.

[36] US Department of Agriculture. Beef 2007-08 Part II: Reference of Beef Cow-calf Management Practices in the United States, 2007-08. Fort Collins, CO: USDA:APHIS:VS, CEAH; 2009. Available from: https://www.aphis.usda.gov/animal_health/nahms/beefcowcalf/downloads/beef0708/Beef0708_dr_PartII.pdf.

[37] US Department of Agriculture. Traceability for Livestock Moving Interstate. Federal Register. 2013;78(6):2040–2075. Available from: http://www.gpo.gov/fdsys/pkg/FR-2013-01-09/pdf/2012-31114.pdf.

[38] Carroll IT, Bansal S. Cattle Auction Markets in the USA: a De-duplicated List Compiled from Trade Association and Federal Agency Directories. V1. Harvard Dataverse. 2014;.

[39] Livestock Marketing Association. Find an Auction/Dealer; 2014. Available from: http://www.lmaweb.com/find-an-auctiondealer-2/.

[40] US Department of Agriculture. Livestock, Poultry, and Grain Market News. Agricultural Marketing Service. 2016;Available from: https://www.ams.usda.gov/AMSv1.0/LPSMarketNewsPage.

[41] Tesseract Development Team. Tesseract OCR. Version 3.04.01; 2016. Available from: https://github.com/tesseract-ocr/tesseract.

[42] FooLabs. Xpdf: A PDF Viewer for X. Version 3.04; 2014. Available from: http://www.foolabs.com/xpdf/home.html.

[43] GeoNames. Web Service Documentation; 2014. Available from: http://www.geonames.org/export/web-services.html.

[44] Oak Ridge National Laboratory. County-to-County Distance Matrix; 2011. Available from: http://cta.ornl.gov/transnet/SkimTree.htm.

[45] Stan Development Team. Stan Modeling Language Users Guide and Reference Manual. Version 2.9.0.; 2016. Available from: http://mc-stan.org.

[46] Chen J, Li H. Variable selection for sparse Dirichlet-multinomial regression with an application to microbiome data analysis. The Annals of Applied Statistics. 2013;7(1):418–442. Available from: http://projecteuclid.org/euclid.aoas/1365527205.

[47] Caswell H. Matrix population models: construction, analysis, and interpretation. Sinauer Associates; 2001. Available from: http://books.google.com/books?id=CPsTAQAAIAAJ.

[48] Mathews KH, Brorsen W, Hahn WF, Arnade C, Dohlman E. Mandatory Price Reporting, Market Efficiency, and Price Discovery in Livestock Markets. Livestock, Dairy, and Poultry Outlook. 2015;LDPM-254-0. Available from: http://www.ers.usda.gov/publications/ldpm-livestock,-dairy,-and-poultry-outlook/ldpm-254-01.aspx.

[49] Feuz DM, Harris C, Bailey D, Halverson G. Transportation and Quality Adjusted Basis: Does the Law of One Price Hold for Feeder Cattle? Journal of Agricultural and Resource Economics. 2008;33(3):498.

[50] Morris M. Network epidemiology: A handbook for survey design and data collection. Oxford University Press; 2004.

[51] Handcock MS, Gile KJ. Modeling social networks from sampled data. The Annals of Applied Statistics. 2010;4(1):5–25. Available from: http://arxiv.org/abs/1010.0891.

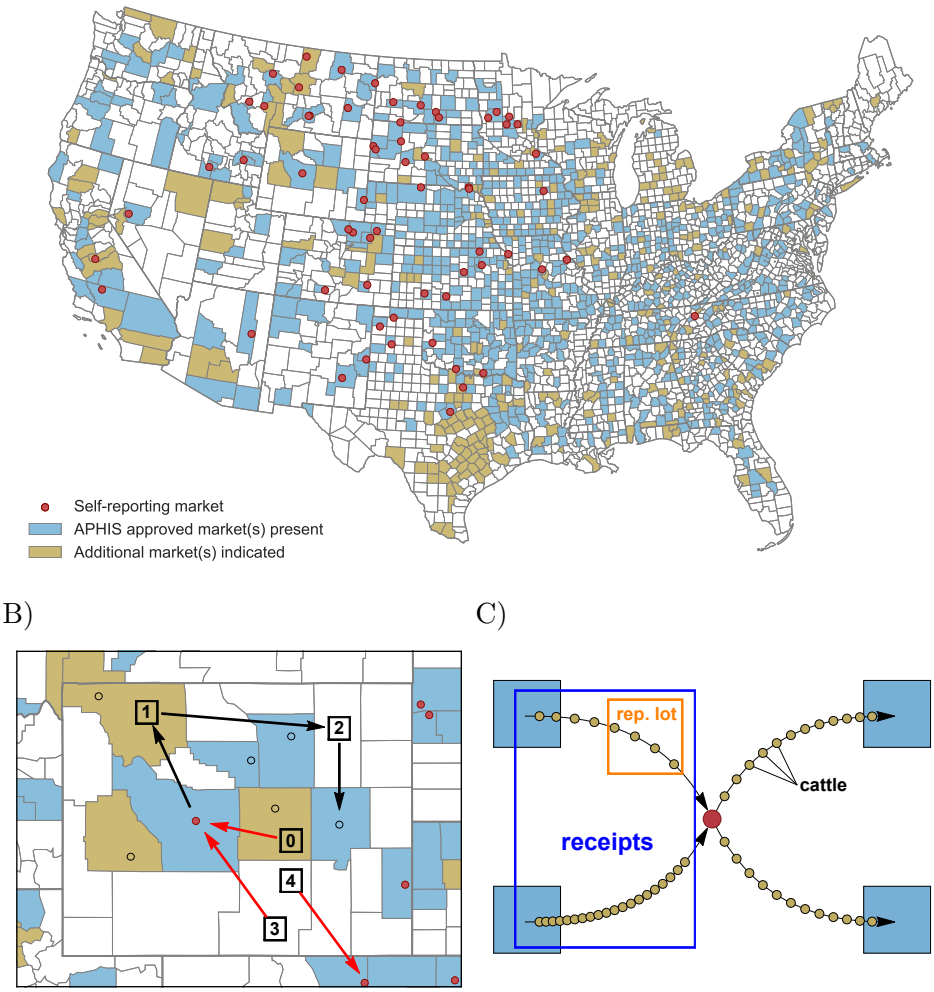
[52] RTI International. GIPSA Livestock and Meat Marketing Study. Research Triangle Park, ND; 2007. Available from: https://www.gipsa.usda.gov/psp/publication/live_meat_market.aspx.

[53] Bates TW, Thurmond MC, Carpenter TE. Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. American Journal of Veterinary Research. 2003;64(2):195–204. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12602589.

[54] Elbakidze L, Highfield L, Ward MP, McCarl Ba, Norby B. Economics Analysis of Mitigation Strategies for FMD Introduction in Highly Concentrated Animal Feeding Regions. Review of Agricultural Economics. 2009;31(4):931–950. Available from: http://aepp.oxfordjournals.org/lookup/doi/10.1111/j.1467-9353.2009.01477.x.

[55] Pomeroy LW, Bansal S, Tildesley M, Moreno-Torres KI, Moritz M, Xiao N, et al. Data-Driven Models of Foot-and-Mouth Disease Dynamics: A Review. Transboundary and Emerging Diseases. 2015 nov;Available from: http://doi.wiley.com/10.1111/tbed.12437.

A)



Box 1. Cattle sales at auction markets are an integral part of the US livestock system. Marketing facilities, in-particular those voluntarily approved by APHIS and subject to ICVI exemptions, are distributed widely throughout the US. Representative lots of cattle sold at markets with online market reports provide a sample of the cattle transportation network. A) Presence of at least one auction market in each US county (blue or gold), or at east one APHIS-approved marketing facility (blue), estimated by compilation of public and private market directories [38]. Auction markets reporting representative lots collected for this study (red points) are concentrated in major beef producing regions. B) Two auction markets in Wyoming advertise representative lots online, revealing movement of cattle from unspecified livestock operations within a known county (numbered squares) to those markets (red arrows). Black arrows represent movement of cattle that is unobserved because the cattle are leaving a market, consigned at a non-reporting market (black circles), or moved directly between livestock producers. Individual animals may pass through multiple markets and farms; for example, a cow-calf operation (0) consigns a weaned steer calf at auction that is purchased by a stocker/back-grounder (1), who has a marketing agreement with a feedlot (2), that consigns the fed steer to an auction market buyer for a meat packing operation (not shown). Our data processing aggregates farms within counties, so farms 3 and 4 are represented in a movement network by a single county node connected to two market nodes. C) Representative lots are small groups of cattle consigned by one operation on the day of sale at a given auction, and receipts are the total number of cattle consigned at that sale. Purchased lots are typically combined with others, but could also be disaggregated, for shipment to the buyer's operation(s).

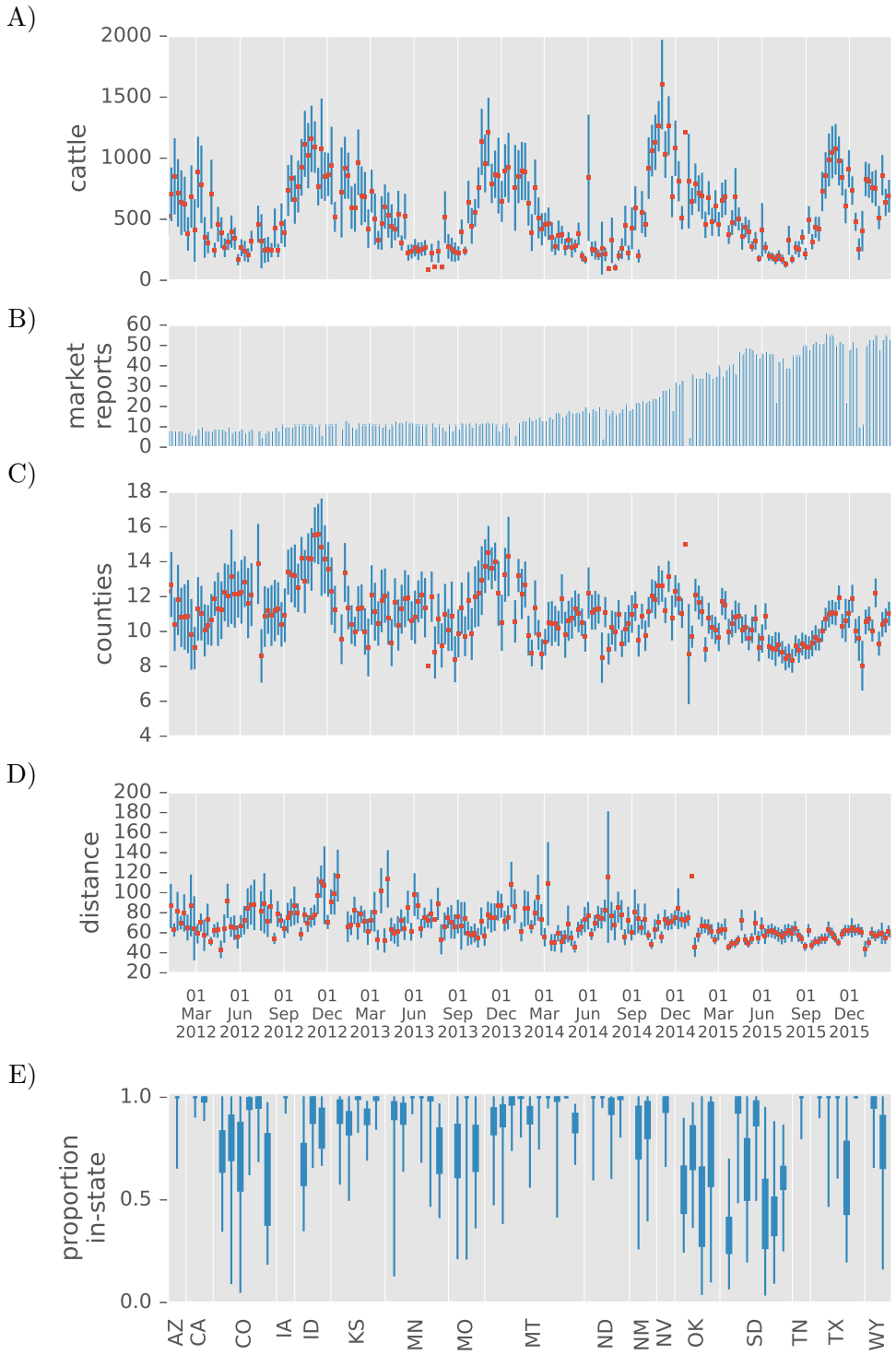


Figure 1. Cattle movements captured in representative lots for a 3.5 year time span from auction markets in the US. A) The total number of reports collected for cattle sales occurring within a given week. B) The average (+/- SEM) head of cattle reported with locations of origin each week by a single livestock market. The outlying observation in the first week of June, 2014 includes a massive auction conducted by the World Livestock Auctioneer Champion in Ft. Pierre, SD. C) The average (+/- SEM) number of counties from which cattle arrived for sale at a market on a given week. D) The average (+/- SEM) distance between county of origin and auction market per head of cattle sold on a given week. E) The proportion of cattle (interquartile and entire ranges) in representative lots transported from a county within the same state to each reporting market.

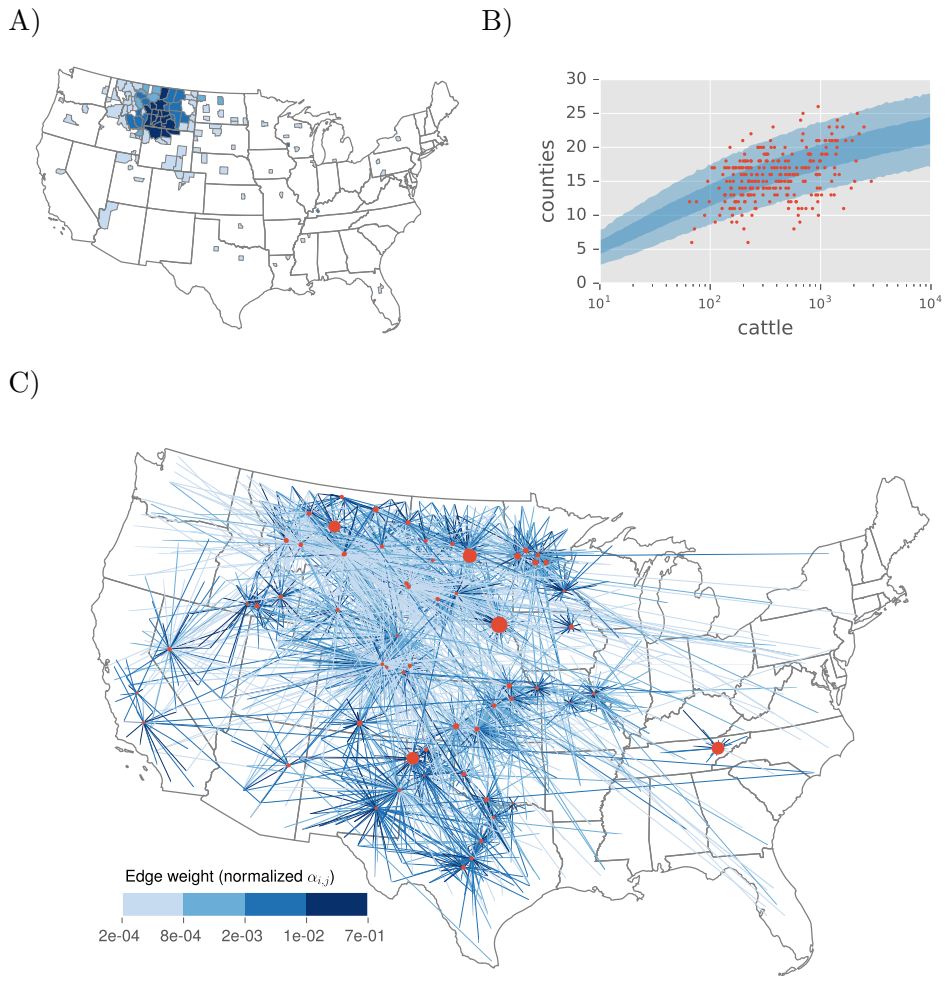


Figure 2. Values of the network model parameters, $\alpha_{i,j}$, that affect the distribution of counties from which cattle are sold when these are fitted directly and independently for each market. Panels A and B exemplify the fit obtained for nearly all markets (Supplementary Figures). A) The posterior mean $\alpha_{i,j}$ for a market i in Montana and counties j from this market's representative lots. B) The predictive distribution (50th and 95th percentiles) on the same market's in-degree for increasing receipts in comparison to individual market reports (points) showing the number of counties and total head within representative lots. C) The posterior mean $\alpha_{i,j}$ for all markets, re-scaled as an edge weight represented by color, along with the log of the total $\alpha_{i,j}$ over all counties j represented as market vertex size.

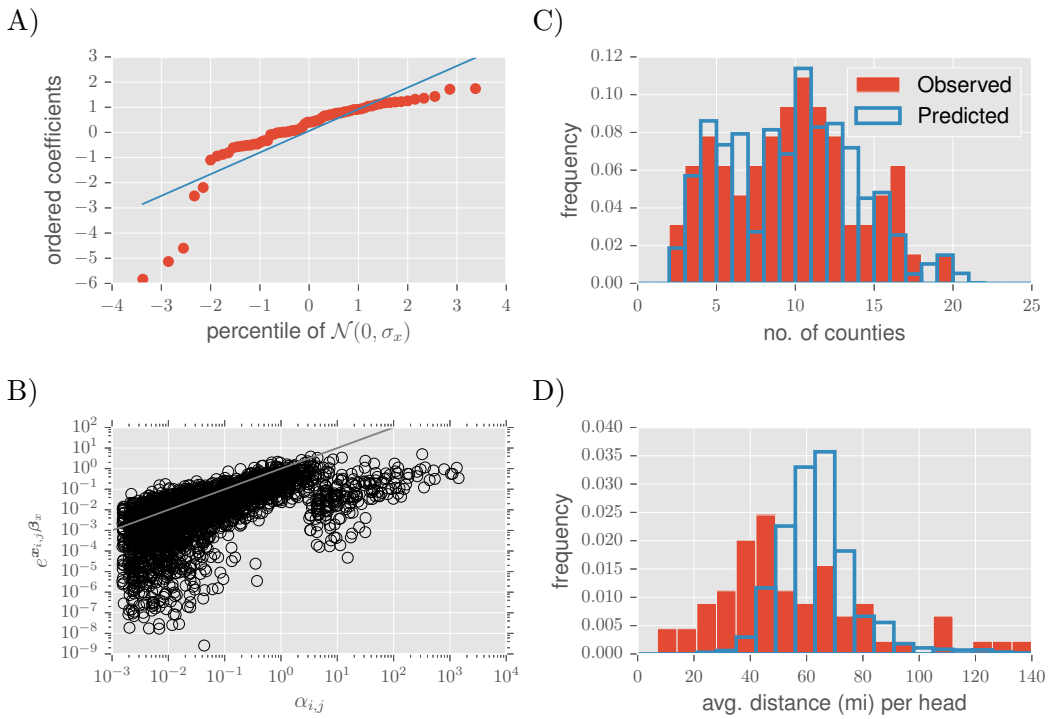


Figure 3. Network model parameters $\boldsymbol{\beta}_x$ affect the distribution of counties sold at each market through linear combination with independent variables $\boldsymbol{x}_{i,j}$. A) Probability plot comparing the posterior mean market intercepts with a normal distribution having SD equal to the posterior mean of σ_x. B) Comparison between the regression model and the saturated model, in which each $\alpha_{i,j}$ is a free parameter. C) The distribution on market average in-degree, or the average number of counties of origin among representative lots for a given auction market. D) The distribution on market average movement distance per head, the distance between centroids of the origin county and the market's county averaged across animals in each market report and subsequently across reports for each market.

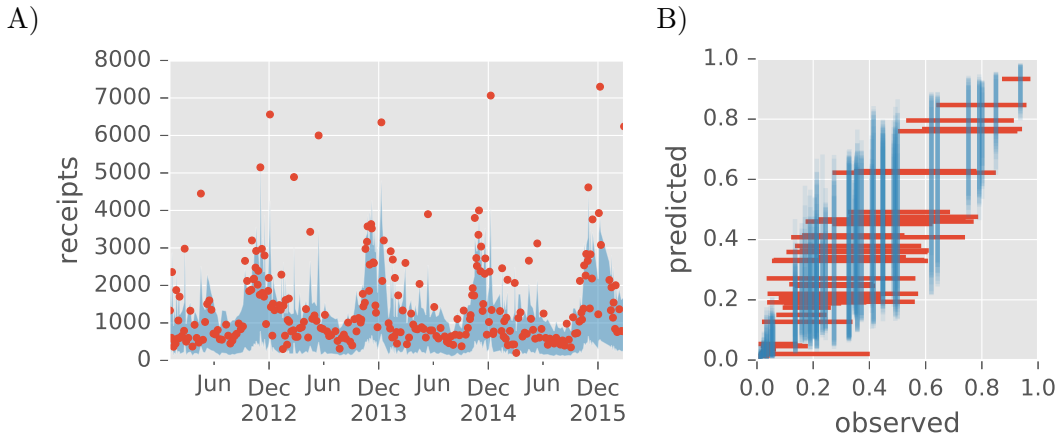


Figure 4. Network model parameters $\boldsymbol{\beta}_y$ affect the number of receipts at each market through linear combination with independent variables $\boldsymbol{y}_{i,j}$. A) The posterior predictive range (95th percentile) on receipts for the Montana market from Fig. 2A with observed receipts (points). B) Within and between market variation in the proportion of receipts included among representative sales. For each market, lines show the inter-95%-tile range of observed (horizontal) and predicted (vertical) proportions intersecting at their respective means.

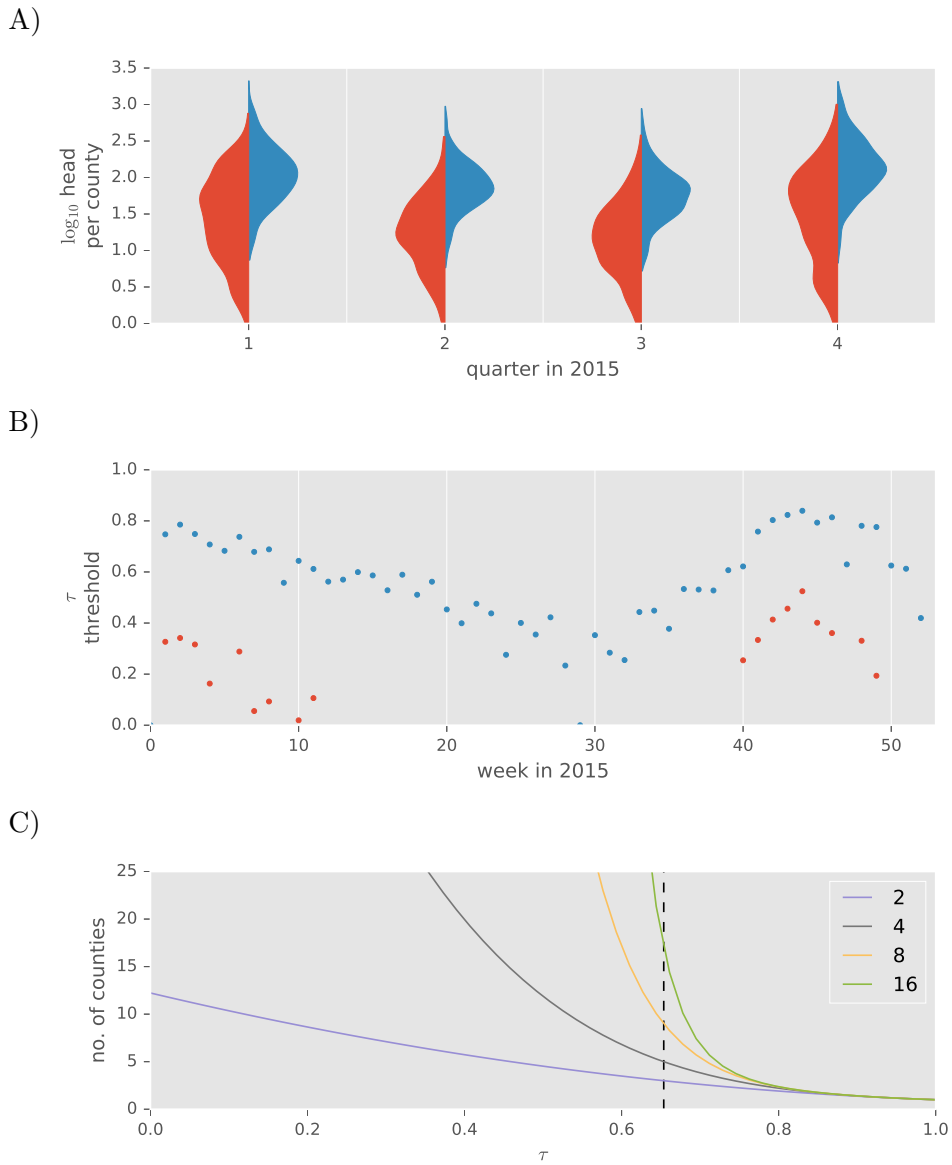


Figure 5. Influence of model parameters on the number of counties designated as control areas following a single introduction of a foreign disease of cattle. A) Observed (red) and modeled (blue) distributions on the the number of cattle shipped to any one market from any one county, aggregated to quarters. B) The time-varying threshold for market biosecurity, τ, that leads to a full-blown epidemic (blue). For comparison, we show results where the numbers of cattle involved in calculating the distribution on Y are estimated from raw averages across representative lots (red). C) The number of counties that become control areas across the range of τ for an outbreak contained after longer delays, interpreted as the maximum number of dangerous contacts a county must back-trace to reach the source of a disease outbreak. Parameter $\lambda = 1$ and $\nu = 0.02$.

A Description of Data Release

Note: The data release will coincide with publication of the article – the DOI given below will remain inactive until release.

Data collected from livestock market websites, along with supporting data from sources attributed below, are available for download from the Bansal Lab Dataverse [1]. The data are separated into four tables that can be joined on one or more of the variables `report`, `market` and `county`. In addition, data on interstate flows [4] are available in electronic form at `http://webarchives.cdlib.org/sw12j6951w/http://www.ers.usda.gov/Data/InterstateLivestockMovements/View.asp`.

---

receipts

report (int, primary key) report identifier

market (int) foreign key linking to the `distance` table

date (char) date string as YYYY-MM-DD of the cattle auction

receipts (int) total number of cattle consigned

mn (char) USDA/AMS Market News (MN) report identifier

mn_receipts (int) receipts from the corresponding MN report

The receipts for each cattle auction as reported by the auction market and as recorded in MN reports [2] for the same sale. Full MN reports are available online using the URL `http://search.ams.usda.gov/mndms/{YYYY}/{MM}/{mn}.TXT` with the substitutions indicated by curly braces.

---

lots

lot (int, primary key) lot identifier

report (int) foreign key linking to receipts

head (int) number of cattle

county (char) County code for the location of origin

The representative lots extracted from each market report with the consignor's location geocoded to county, and the number of cattle consigned together (i.e. as a lot). The `report` identifier provides a link to additional information on each report in `receipts`.

---

distance

market (int, primary key) market identifier

county (char, primary key) County codes used in this analysis

distance (decimal) miles along the great circle connecting centroids of this county and this market's county

Distances are reproduced for convenience from [5]. The county containing a particular market is the one with zero distance.

---

census

county (char, primary key) County codes used in this analysis

inv_num_cattle (decimal) Inventory of cattle and calves

inv_num_other (decimal) Inventory of other cattle and claves

inv_num_cows (decimal) Inventory of cows and heifers that calved

inv_num_beef_cows (decimal) Inventory of beef cows

inv_num_milk_cows (decimal) Inventory of beef cows

inv_num_on_feed (decimal) Inventory of cattle on feed

sale_num_cattle (decimal) Sales by number of cattle

sale_num_calves (decimal) Sales by number of calves

sale_num_500lbs (decimal) Sales by number of cattle weighing over 500lbs

sale_num_on_feed (decimal) Sales by number of cattle on feed

These data are reproduced for convenience from [3] and are a selection of variables for the year 2012 census as seen in the county tables titled "Cattle and Calves – Inventory and Sales' for each state (see http://www.agcensus.usda.gov/Publications/2012/ for PDF format and census methodology). The following relationships hold in principle:

$$\begin{aligned} \texttt{inv_num_cattle} &= \texttt{inv_num_cows} + \texttt{inv_num_other}, \\ \texttt{inv_num_cows} &= \texttt{inv_num_beef_cows} + \texttt{inv_num_milk_cows}, \text{ and} \\ \texttt{sale_num_cattle} &= \texttt{sale_num_calves} + \texttt{sale_num_500lbs}. \end{aligned}$$

Decimal values (rather than integer numbers of cattle) are possible due to the process of approximating missing data, which is described in the main text.

References

[1] Carroll, IT, Bansal S. Replication Data for: Livestock market data for modeling disease spread among US cattle. Harvard Dataverse 2015. DOI: http://dx.doi.org/10.7910/DVN/MAXZAY.

[2] USDA\AMS. Livestock, Poultry, and Grain Market News;. Available from: http://www.ams.usda.gov/AMSv1.0/LPSMarketNewsPage.

[3] USDA\NASS. 2012 Census of Agriculture; 2015. Available from: ftp://ftp.nass.usda.gov/quickstats/qs.census2012.txt.gz.

[4] Shields DA, Mathews KH. Interstate Livestock Movements. USDA:ERS; 2003. Available from: http://www.ers.usda.gov/media/312234/ldpm10801_1_.pdf.

[5] County-to-County Distance Matrix; 2011. Available from: http://cta.ornl.gov/transnet/SkimTree.htm.

B Supplementary Methods

Branching Process Model

The random variable Y appearing in equation 4 represents the number of "children" for any "parent" in a branching process. Given a distribution on Y, it is possible to calculate the distribution on the size of the entire tree, although we only consider its first two moments. In the livestock disease scenario described in the main text, Y is the number of control areas which arise because a farm in a new area received cattle from an area that will also be designated a control area. Farms within a designated control area have all movements restricted, but there is a window between the first dangerous contact and restriction of all movements during which consignment of cattle at auction can induce further dangerous contacts.

We need several assumptions in determining a suitable distribution on Y, some to stand in for processes unrelated to any available data, such as the purchasing of auctioned cattle. We assume that the number of sales, M, at which operations from a given control area consign cattle during the time between the first dangerous contact by a farm within that area and the implementation of a movement ban has a Poisson distribution with mean λ. We denote by N_i the number of cattle consigned at sale i of M by all the farms from an area soon to be controlled, and use market report data to estimate the distribution on N_i as described in the main text. Finally, we assume that operations purchasing one or more of the N_i cattle are distributed among C_i unique areas, so

$$Y = \sum_{i=1}^{M} C_i. \tag{B.1}$$

By independence among the C_i, we have the following conditional expectation

$$\langle z^Y | M, N_1, N_2, \ldots, N_M \rangle = \prod_{i=1}^{M} \langle z^{C_i} | N_i \rangle. \tag{B.2}$$

For the distribution on C_i, we suppose a zero-inflation parameter τ gives the probability that the market hosting sale i quarantines all cattle from an area that will become controlled. Additionally, we suppose a splitting parameter ν that puts all buyers in the area where the cattle originate when $\nu = 0$ and sends each animal to a unique area when $\nu = 1$. In particular, we assume a binomial distribution on the range 0 to N_i for the number of receiving areas when cattle are not quarantined, giving

$$\langle z^{C_i} | N_i \rangle = \tau + (1-\tau) \sum_{j=0}^{N_i} z^j \binom{N_i}{j} \nu^j (1-\nu)^{N_i - j}. \tag{B.3}$$

After averaging over the identically distributed N_i, to obtain $\langle z^C \rangle$, and employing the generating function for the Poisson distribution,

$$G(z) = \langle z^Y \rangle = e^{\lambda\left(\langle z^C \rangle - 1\right)}, \tag{B.4}$$

is the generating function for Y. From it, we obtain the first two factorial moments

$$\langle Y \rangle = \lambda(1-\tau)\nu\langle N \rangle \tag{B.5}$$

$$\langle Y(Y-1) \rangle = \lambda(1-\tau)\nu^2 \left(\lambda(1-\tau)\langle N \rangle^2 + \langle N(N-1) \rangle \right). \tag{B.6}$$

A generating function for the distribution on the total size of the branching process, or total number of control areas eventually established, is the obtained by application of the self-consistency relations described by Newman [1] for epidemics spreading on a random graph. We add the minor variation of keeping track of "generations", so with random variable X_n the size of a tree descending no more than n generations from a common ancestor, we impose the inter-generational consistency relation

$$X_0 = 1 \tag{B.7}$$

$$X_{n+1} = 1 + \sum_{i=1}^{Y} X_n. \tag{B.8}$$

Because every X_n in the sum is independent

$$\langle z^{X_{n+1}} | Y \rangle = z \langle z^{X_n} \rangle^Y. \tag{B.9}$$

Denoting the generating function that has appeared on the right as $H_n(z)$, we obtain

$$H_{n+1}(z) = zG(H_n(z)). \tag{B.10}$$

In theory, the number of control areas that must eventually be established depends on $H_n(z)$ as $n \to \infty$, but the premise that Y does not tend to diminish over the course of an outbreak should not be taken too seriously. For an outbreak that is controlled within a small number of generations, however, the assumption of a stationary Y is a reasonable approximation.

Stan Model Code

The following code in the Stan modelling language provides instructions for MCMC sampling of parameters affecting the likelihood on the number of cattle observed in representative lots consigned to a given market from different counties. Our likelihood function is the Pólya probability distribution parameterized by α_j for j one of c counties (the subscript for market is dropped in this section because independent fits are performed for each market). The code implements a novel approach to calculating the likelihood for independent samples, i.e. the $n_{j,i}$ head for each i of s reports from a focal market, drawn from a Pólya distribution, which would be naively calculated as:

$$\prod_{i=1}^{s} \frac{\Gamma(\alpha_0)\Gamma(n_{0,i}+1)}{\Gamma(\alpha_0 + n_{0,i})} \prod_{j=1}^{m} \frac{\Gamma(\alpha_j + n_{j,i})}{\Gamma(\alpha_j)\Gamma(n_{j,i}+1)}, \tag{B.11}$$

where $\alpha_0 = \sum_{j=1}^{m} \alpha_j$ and $n_{0,i} = \sum_{j=1}^{m} n_{j,i}$ [2]. Calculating the log likelilood this way requires a routine for the log-gamma function, which must be called at least $m+1$ times *for each sample*.

Calculating the log likelihood with an approach having constant complexity with respect to the number of samples is made possible by the identity

$$\ln \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)} = \sum_{k=0}^{n-1} \ln(\alpha + k), \forall\, n \in \mathbb{Z}^+. \tag{B.12}$$

For the new approach, we first expand each log-gamma function according to this identity while using the Boolean function, $\mathcal{B}(\cdot)$, to substitute for the summation limit:

$$\sum_{i=1}^{n} \sum_{k=0}^{\infty} -\ln(\alpha_0 + k)\mathcal{B}(n_{0,i} > k) + \sum_{j=1}^{m} \ln(\alpha_j + k)\mathcal{B}(n_{j,i} > k), \tag{B.13}$$

ignoring terms constant with respect to every α_j. The sum across samples can now take place before parameter estimation begins, and the log likelihood is calculated as

$$\sum_{j=1}^{m}\sum_{k=0}^{\infty}\ln(\alpha_j + k)y_{j,k}, \text{ where} \tag{B.14}$$

$$y_{j,k} = \sum_{i=1}^{s}\mathcal{B}(n_{j,i} > k), \text{ and} \tag{B.15}$$

$$y_{0,k} = -\sum_{i=1}^{s}\mathcal{B}(n_{0,i} > k). \tag{B.16}$$

The double summation is implemented as a dot product in the code below, with sparse matrix $\boldsymbol{y}_{\cdot,\cdot}$ provided as a coordinate list (sparse COO storage) and with values of $\ln(\alpha_j + k)$ calculated as needed.

```
data {
    /* Dependent variable */
    int N;  // number of non-zero values in sparse matrix
    int M;  // number of counties with observations
    vector[N] y;    // response value, in coordinate list (COO) sparse
    vector[N] y_i;  // matrix format, such that y[k] reports have more
    int y_j[N];     // than y_i[k] animals from county y_j[k]
}
parameters {
    vector<lower=0>[M] alpha;
}
model {
    /* Prior */
    alpha ~ cauchy(0, 5);

    /* Likelihood */
    {
        vector[M + 1] x;
        vector[N] z;

        x[1:M] <- alpha;
        x[M + 1] <- sum(alpha);
        z <- log(x[y_j] + y_i);
        increment_log_prob(dot_product(z, y));
    }
}
```

The next code block specifies a multi-level, generalized linear model with log-link function and Pólya likelihood. The response variable is transformed as above into a sparse COO format, with a row for each livestock market and county pair. The hierarchical term adds a random intercept for each livestock market. The data under `/* Other */` provides indexing to correctly form and place $\sum_{j=1}^{c}\alpha_{i,j}$ for each market i for use in the likelihood calculation.

```
data {
    /* Dependent variable */
    int N;  // number of samples
    int Q;  // number of non-zero values in sparse sample matrix
```

```
    vector[Q] y;    // response value, in coordinate list (COO) sparse
    int y_i[Q];     // matrix format, such that y[k] reports have more
    vector[Q] y_j;  // than y_j[k] animals from market:county y_i[k]

    /* Categorical independent variables */
    int K;  // number of intercepts
    int L;  // number of variables
    int C[L];  // position of last coefficient in each category
    int M;  // number of non-zero values in sparse design matrix for intercepts
    vector[M] c_w;   // design matrix in compressed row storage where c_w[i]
    int c_v[M];      // is a value from column c_v[i], row j values begin
    int c_u[N + 1];  // at position c_u[j] in c_w, and c_u[N + 1] = M.

    /* Numerical independent variables */
    int P;  // number of slopes
    matrix[N, P] x;  // variables

    /* Other */
    int R_1;
    int R_2;
    int b_1[R_1];
    int b_1_idx[R_1];
    int b_2[R_2 + 1];
    int b_2_idx[R_2];
}
parameters {
    vector[C[L - 1]] alpha_1;  // fixed intercepts
    vector[C[L] - C[L - 1]] alpha_2_raw;  // random intercepts
    real<lower=0> sigma_2;  // s.d. of random intercepts
    vector[P] beta;  // fixed slopes
}
model {
    /* Prior */
    append_row(alpha_1, beta) ~ normal(0, 100);
    alpha_2_raw ~ normal(0, 1);
    sigma_2 ~ cauchy(0, 5);

    /* Likelihood */
    {
        vector[K] alpha;
        vector[N] a;
        vector[R_1 + R_2] b;

        alpha <- append_row(alpha_1, alpha_2_raw * sigma_2);
        a <- csr_matrix_times_vector(N, K, c_w, c_v, c_u, alpha) + x * beta;
        a <- exp(a);
        b[b_1_idx] <- a[b_1];
        for (i in 1:R_2) {
            b[b_2_idx[i]] <- sum(a[b_2[i]:(b_2[i+1] - 1)]);
        }
        increment_log_prob(dot_product(log(b[y_i] + y_j), y));
    }
}
```

The last code block specifies a multi-level, generalized linear model with log-link function

and negative binomial likelihood. The hierarchical term again provides a random intercept for each livestock market. Each record corresponds to one livestock auction with response variable equal to receipts minus the total number of cattle in representative lots.

```
data {
    /* Dependent variable */
    int N;  // number of samples
    int y[N];  // receipts - total of representative lots

    /* Categorical independent variables */
    int K;  // number of intercepts
    int L;  // number of variables
    int C[L];  // position of last coefficient per variable
    int M;  // number of non-zero values in design matrix
    vector[M] d_w;   // design matrix in compressed row storage where d_w[i]
    int d_v[M];      // is a value from column d_v[i], row j values begin
    int d_u[N + 1];  // at position d_u[j] in d_w, and d_u[N + 1] = M

    /* Numerical independent variables */
    int P;  // number of slopes
    matrix[N, P] x;  // variables
}
parameters {
    vector[C[L - 1]] alpha_1;  // fixed intercepts
    vector[C[L] - C[L - 1]] alpha_2_raw;  // random intercepts
    real<lower=0> sigma_2;  // s.d. of random intercepts
    vector[P] beta;  // fixed slopes
    real<lower=0> phi;
}
model {
    /* Prior */
    phi ~ cauchy(0, 5);
    append_row(alpha_1, beta) ~ normal(0, 100);
    alpha_2_raw ~ normal(0, 1);
    sigma_2 ~ cauchy(0, 5);

    /* Likelihood */
    {
        vector[N] log_mu;
        vector[K] alpha;

        alpha <- append_row(alpha_1, alpha_2_raw * sigma_2);
        log_mu <- csr_matrix_times_vector(N, K, d_w, d_v, d_u, alpha) + x * beta;
        y ~ neg_binomial_2_log(log_mu, phi);
    }
}
```

References

[1] Newman, M. E. J.. 2002. Spread of epidemic disease on networks. Physical Review E 66(1): 016128. DOI: 10.1103/PhysRevE.66.016128.

[2] Chen, J. and Li, H.. 2013. Variable selection for sparse Dirichlet-multinomial regression with an application to microbiome data analysis. The Annals of Applied Statistics 7(1):418–442. DOI: 10.1214/12-AOAS592.

C Supplementary Figures

A)

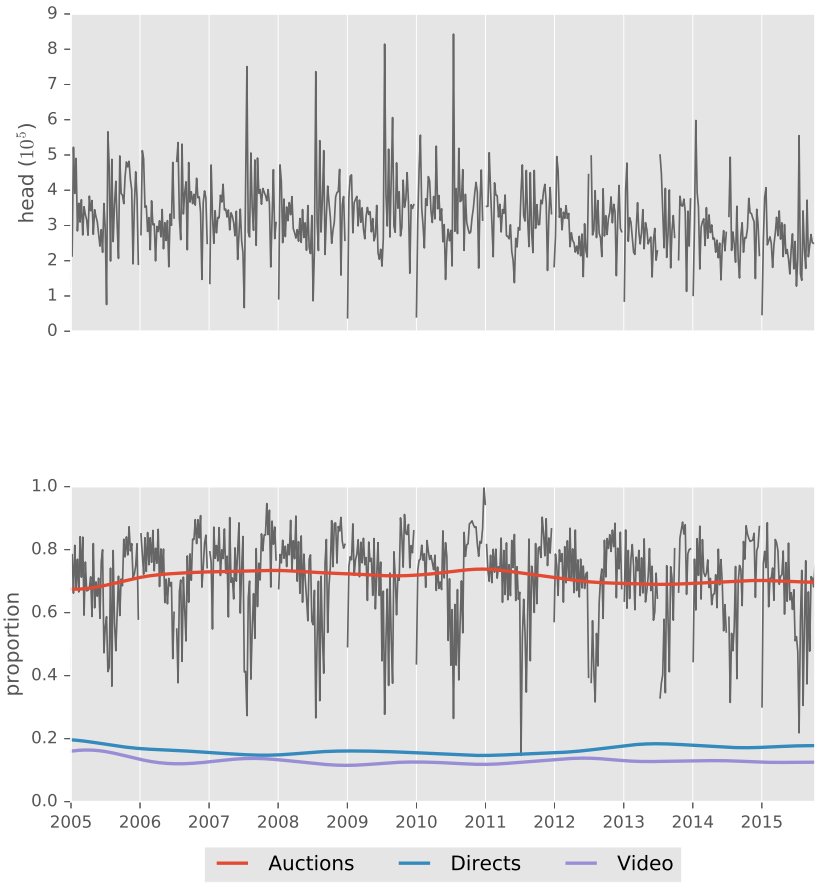
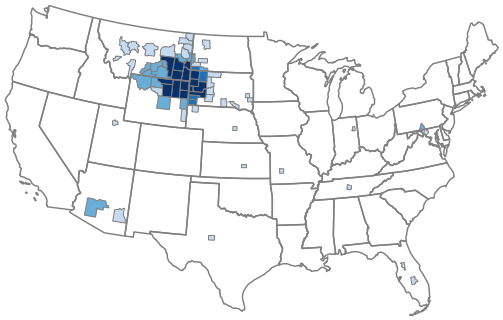
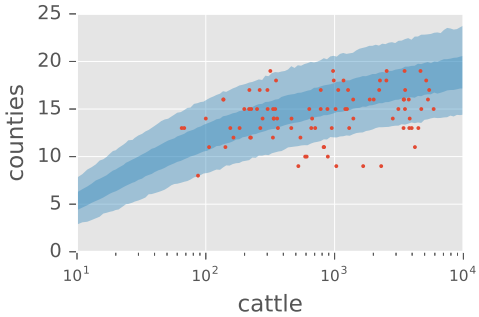


Figure C.1: Feeder and stocker cattle, which are more epidemiologically relevant than cattle sold directly to meat packing operations, are predominantly sold through auction markets. A) Weekly volumes of feeder and stocker cattle sold. B) Proportion sold through auction markets (gray), and smoothed proportions for auction markets along with the next most common marketing arrangements. In contrast, the proportion of slaughter cattle sold through auction markets has declined in recent years [1]. National Feeder and Stocker Cattle Summary data were provided by USDA AMS.

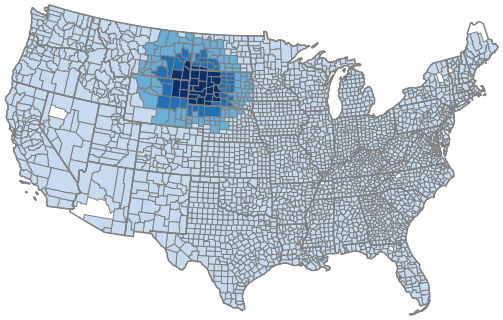
1.A)



1.B)

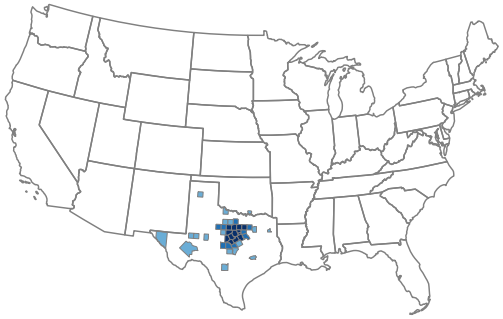


1.C)

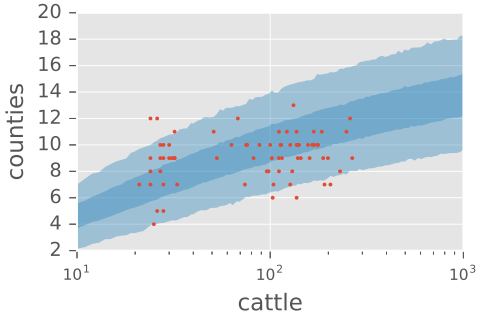


1.D)

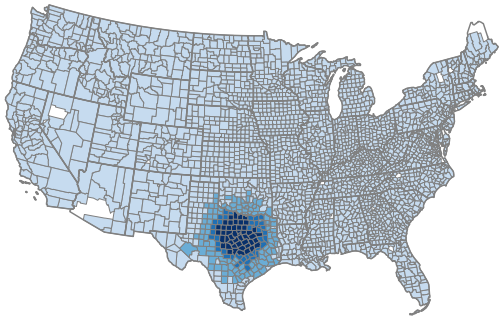
2.A)



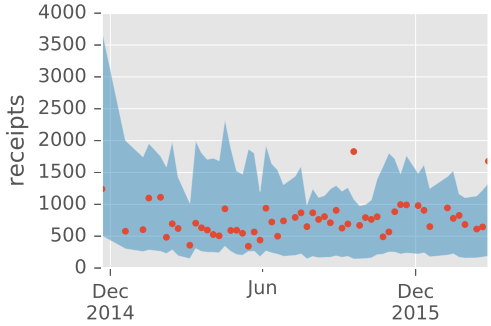
2.B)



2.C)

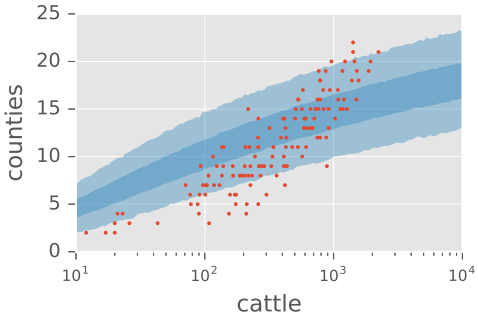


2.D)

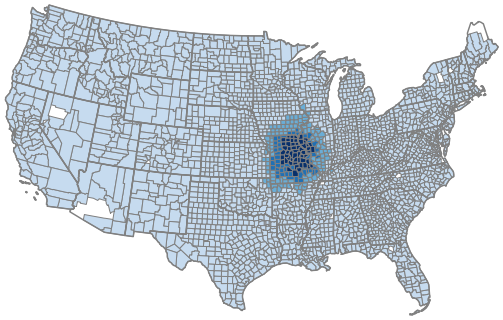


3.A)

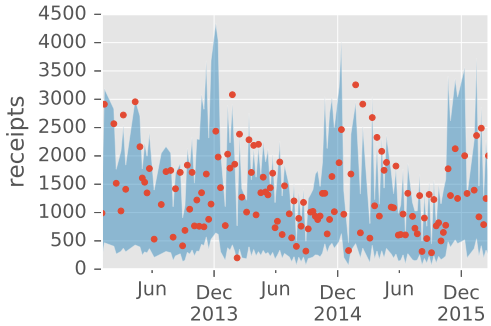
3.B)



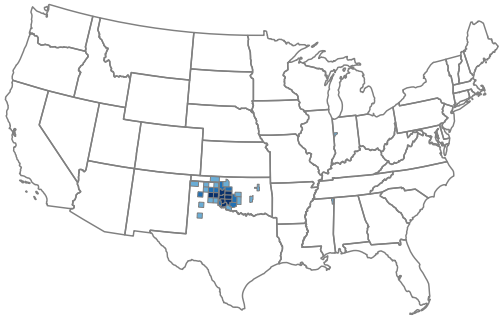
3.C)



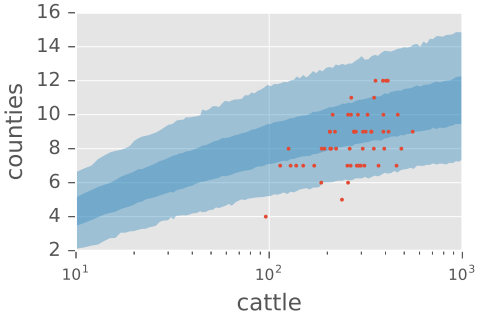
3.D)



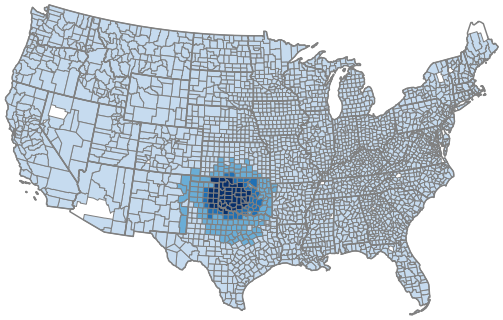
4.A)



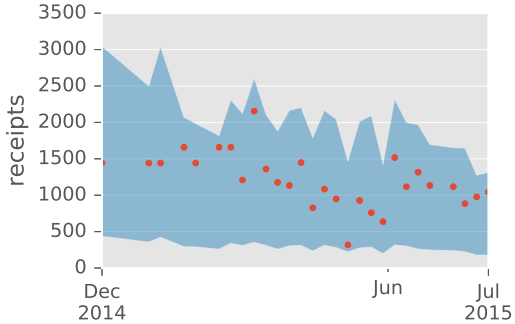
4.B)



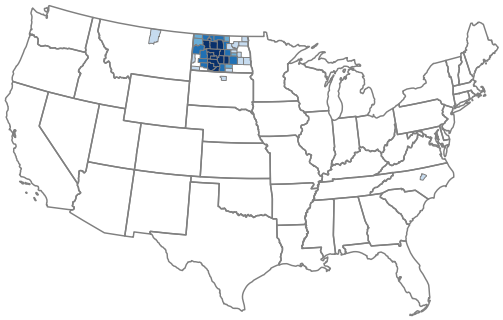
4.C)



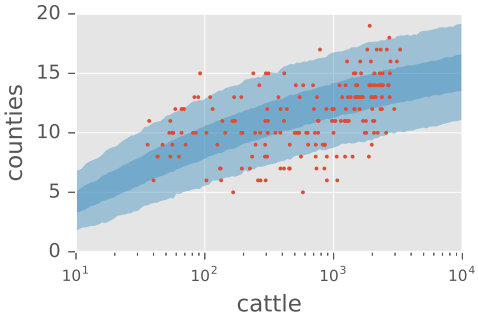
4.D)



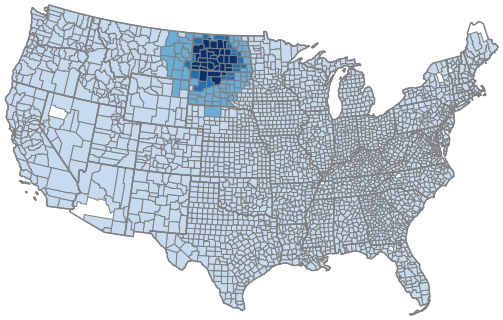
5.A)



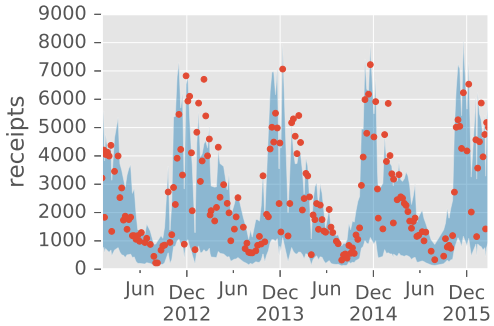
5.B)



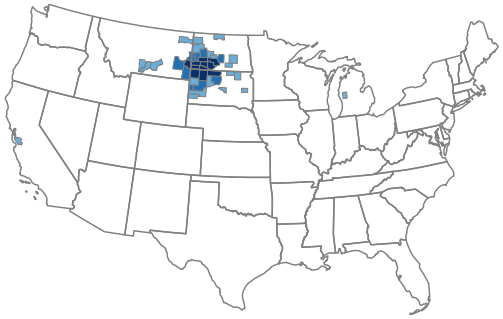
5.C)



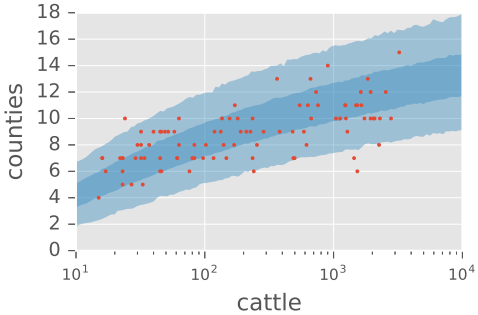
5.D)



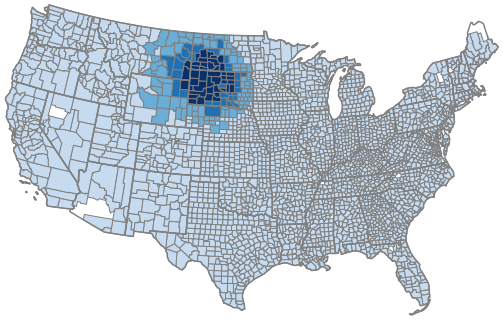
6.A)



6.B)

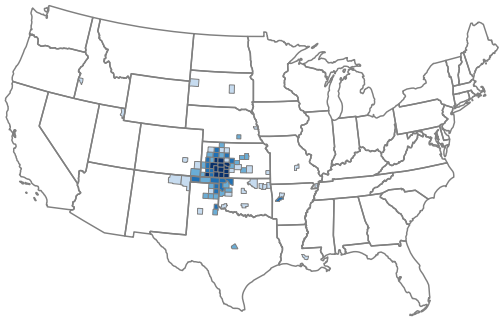


6.C)

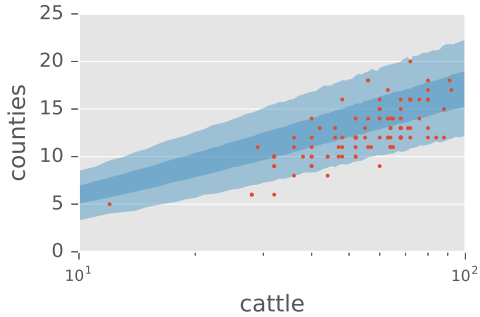


6.D)

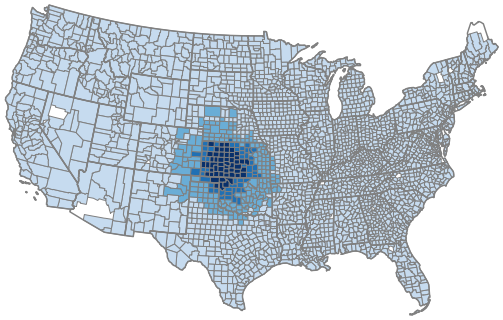
7.A)



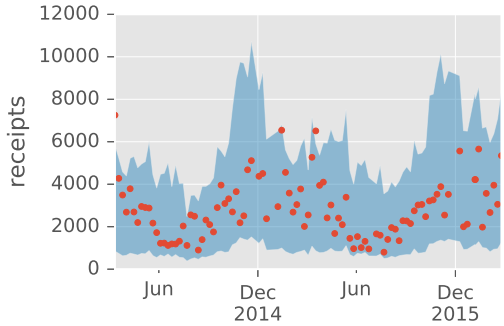
7.B)



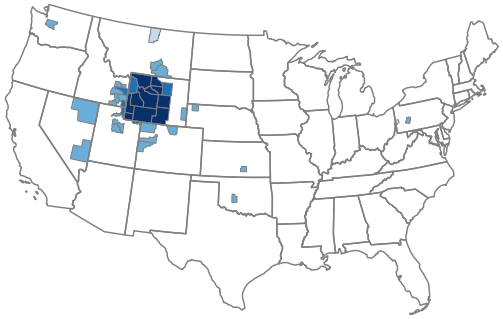
7.C)



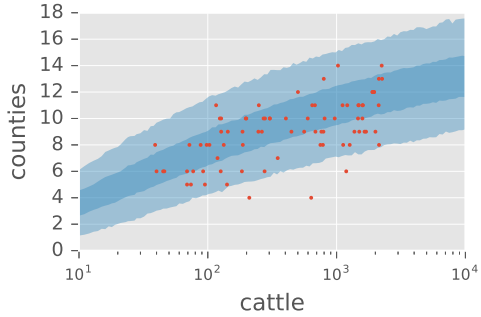
7.D)



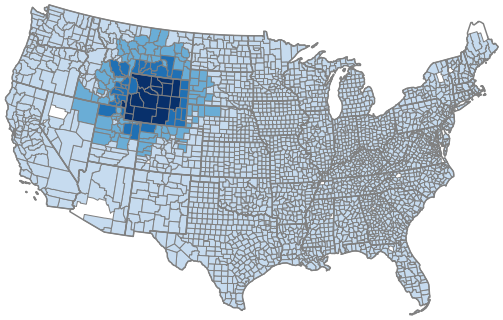
8.A)



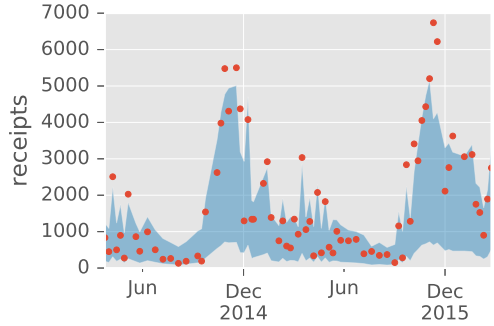
8.B)



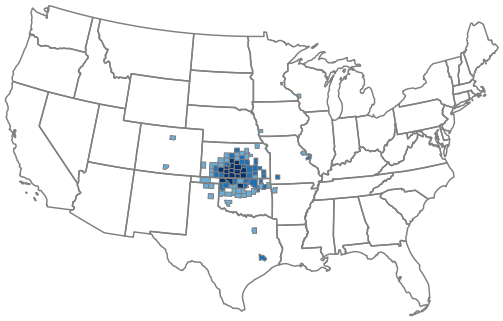
8.C)



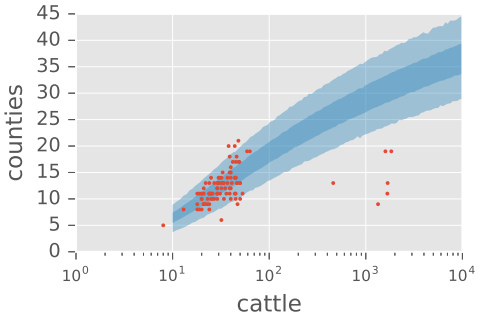
8.D)



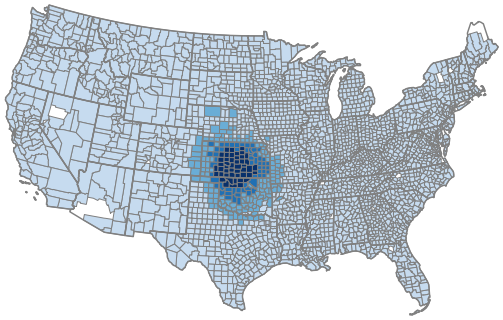
9.A)



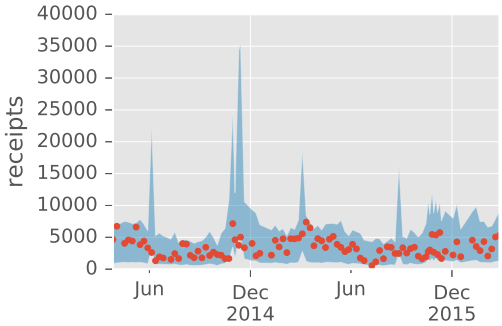
9.B)



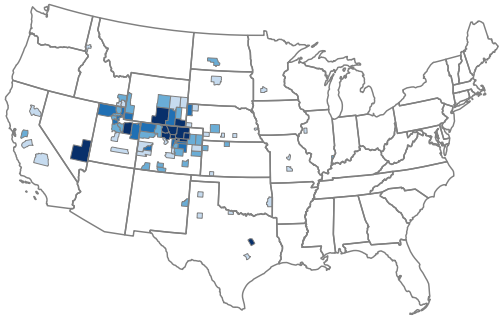
9.C)



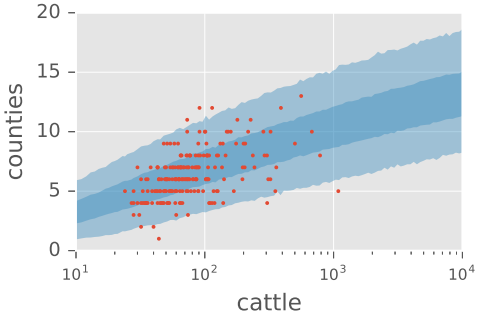
9.D)



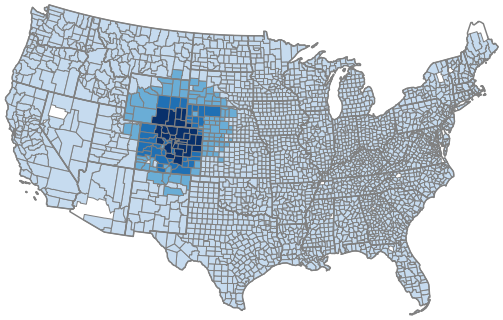
10.A)



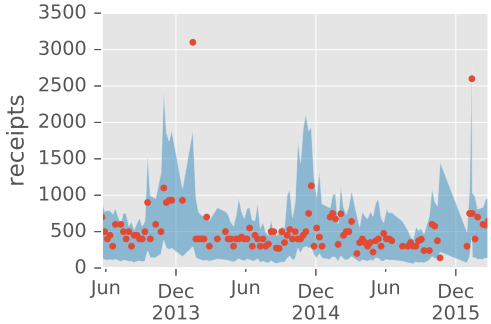
10.B)



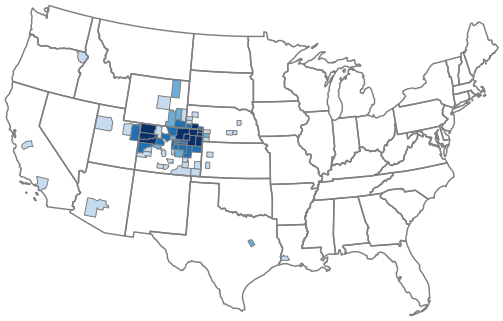
10.C)



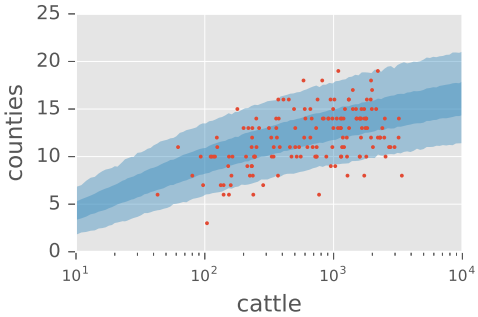
10.D)



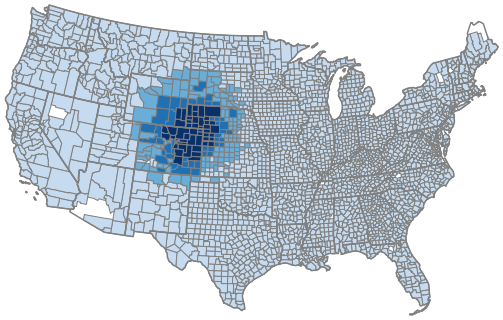
11.A)



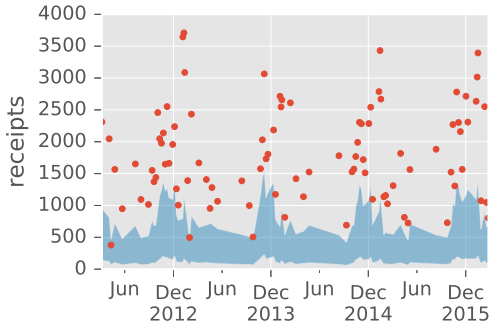
11.B)



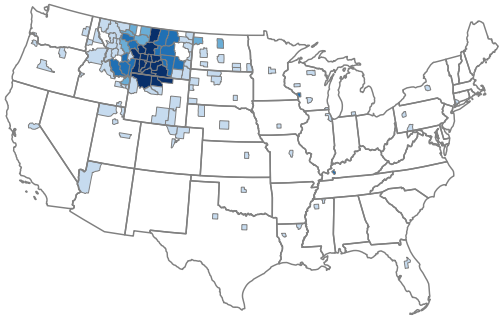
11.C)



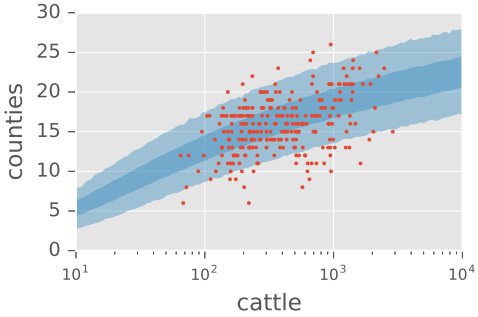
11.D)



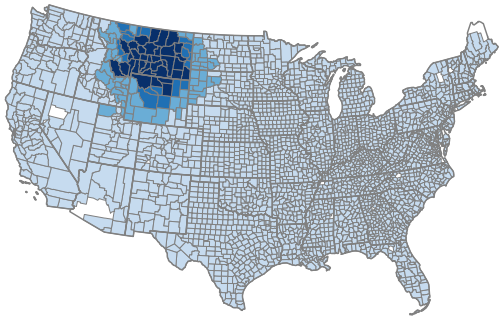
12.A)



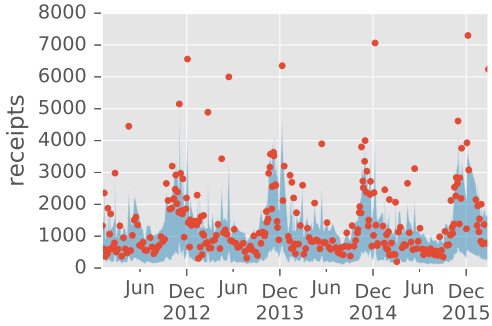
12.B)



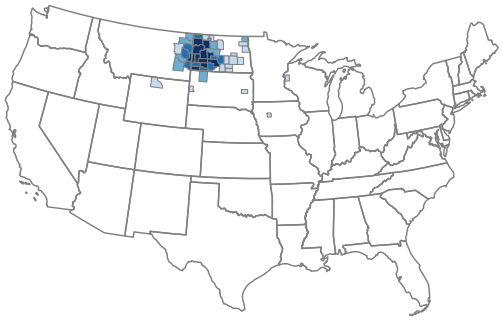
12.C)



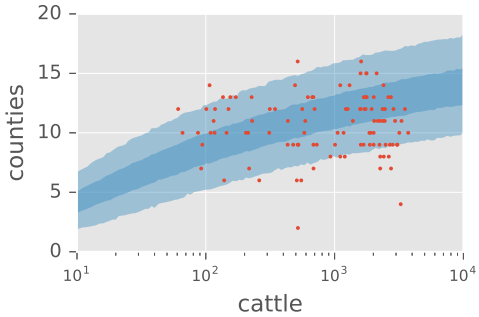
12.D)



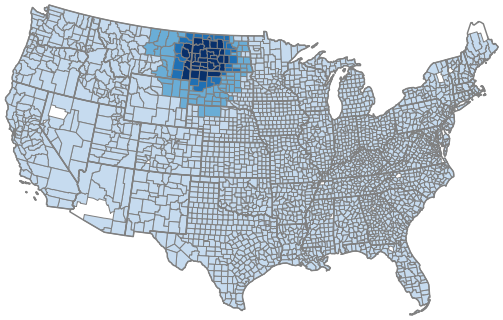
13.A)



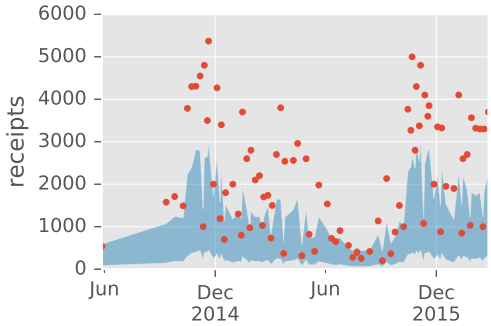
13.B)



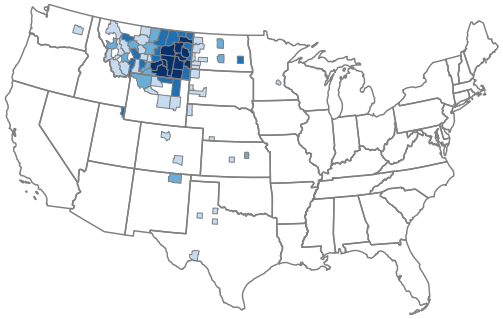
13.C)



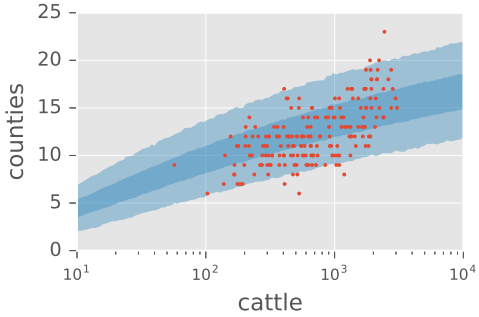
13.D)



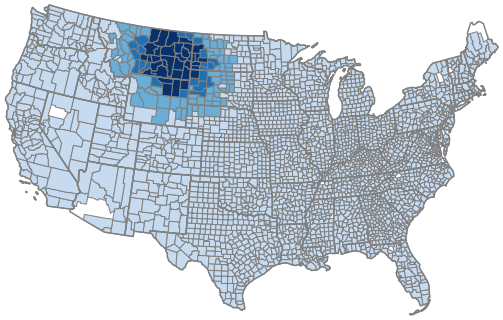
14.A)



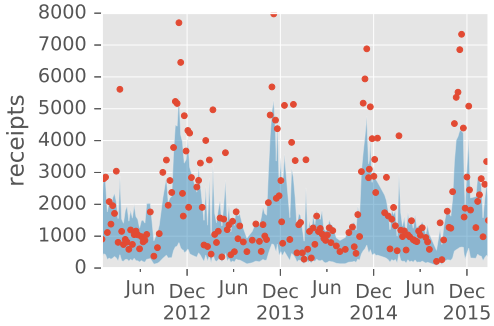
14.B)



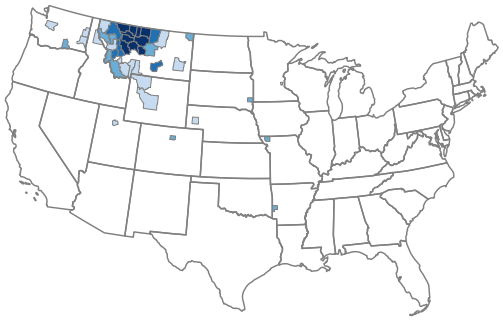
14.C)



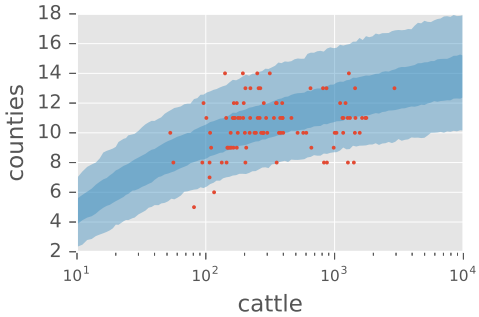
14.D)



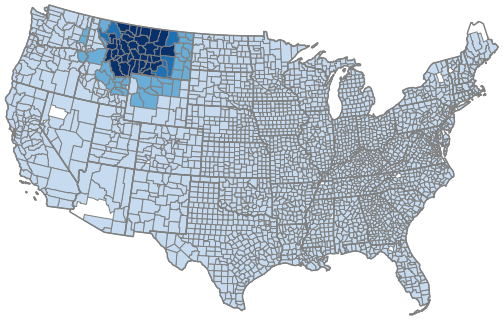
15.A)



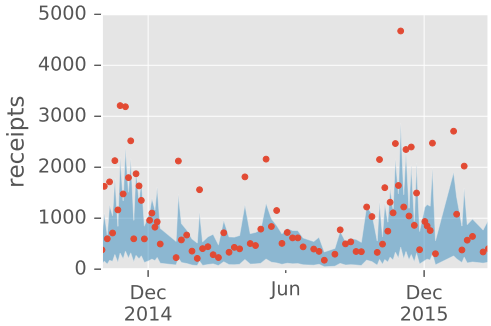
15.B)



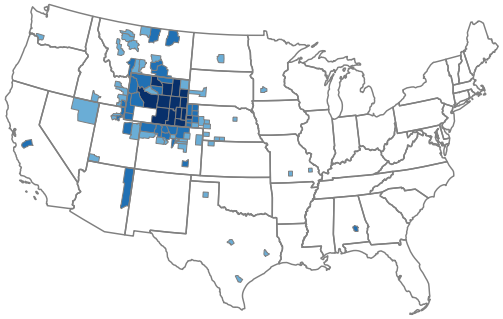
15.C)



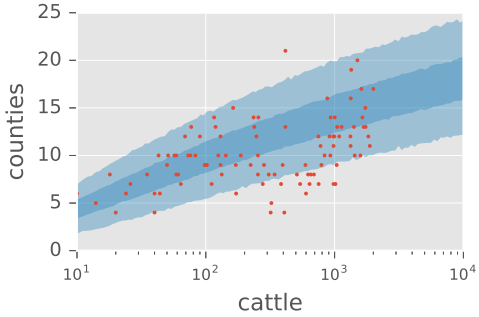
15.D)



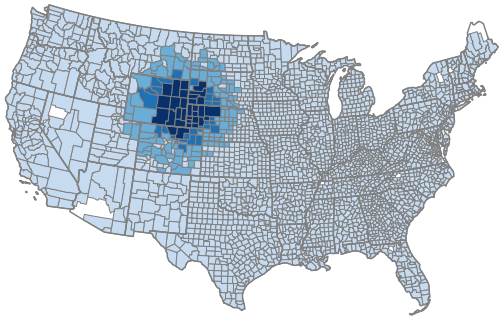
16.A)



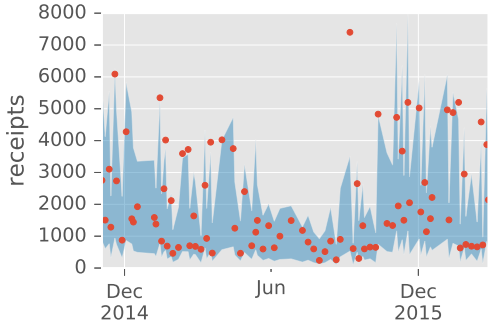
16.B)



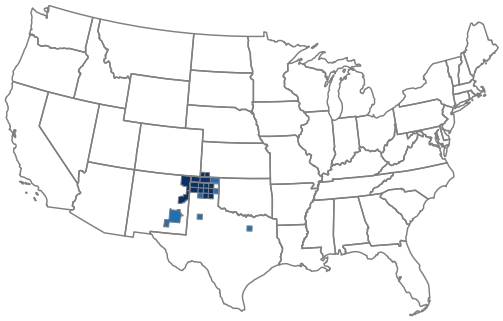
16.C)



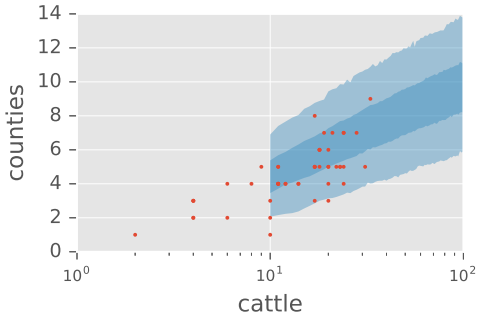
16.D)



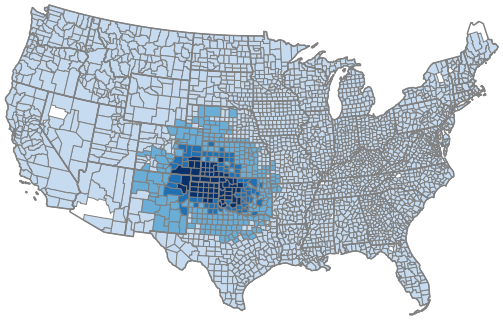
17.A)



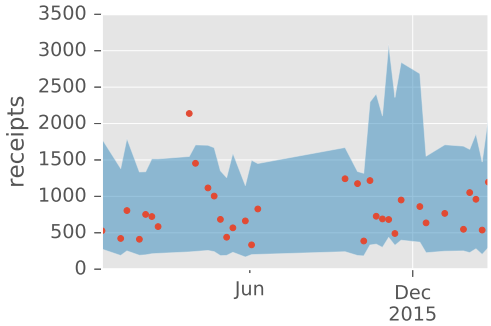
17.B)



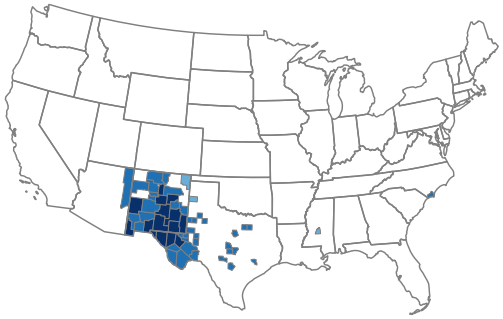
17.C)



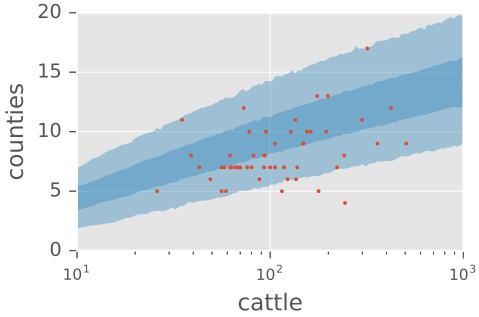
17.D)



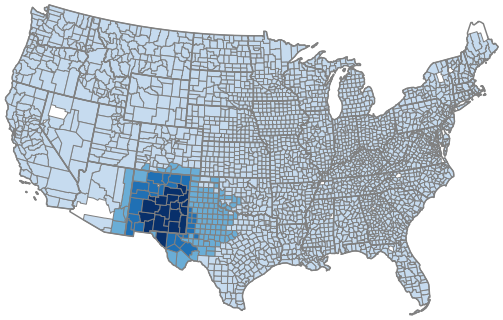
18.A)



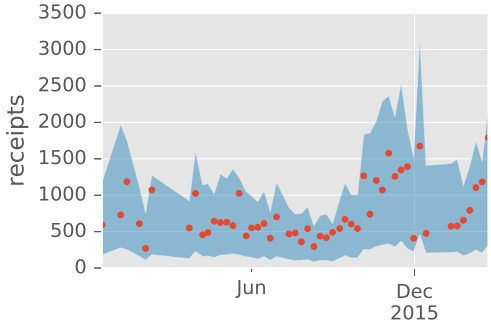
18.B)



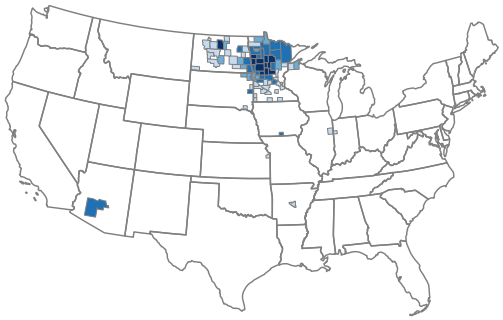
18.C)



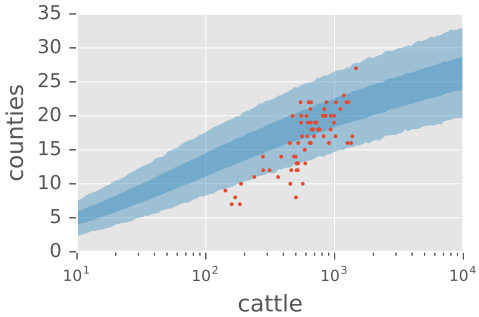
18.D)



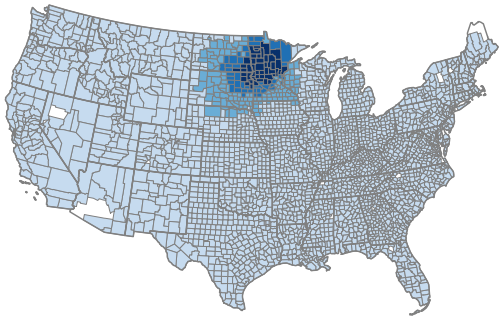
19.A)



19.B)

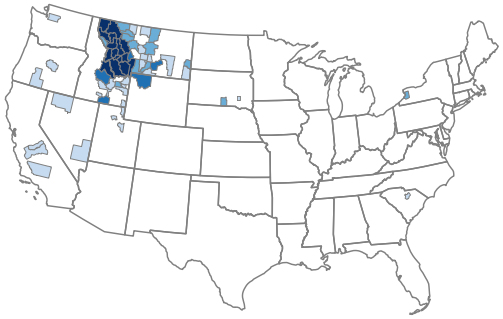


19.C)

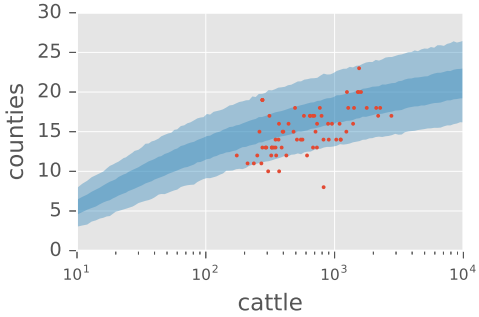


19.D)

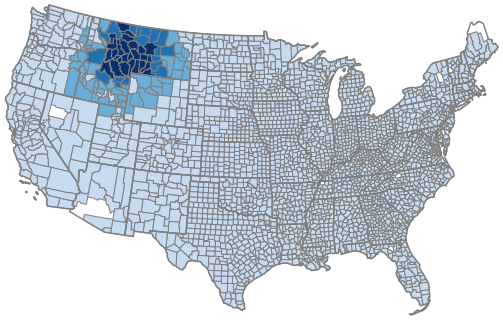
20.A)



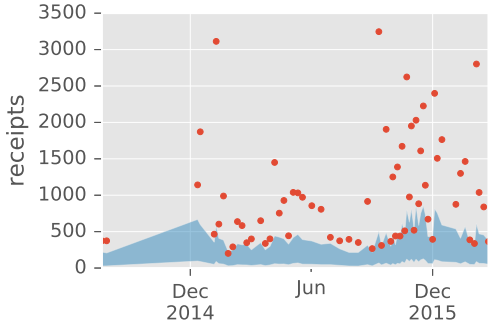
20.B)



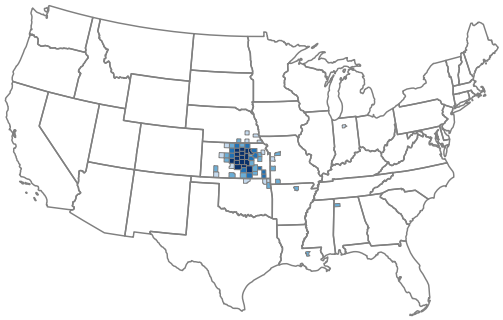
20.C)



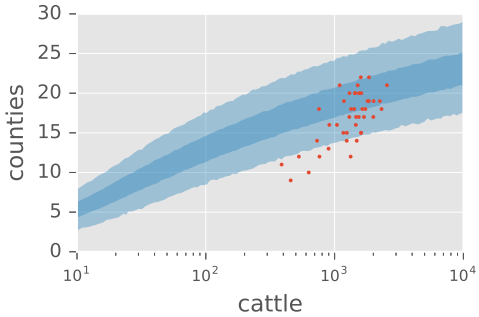
20.D)



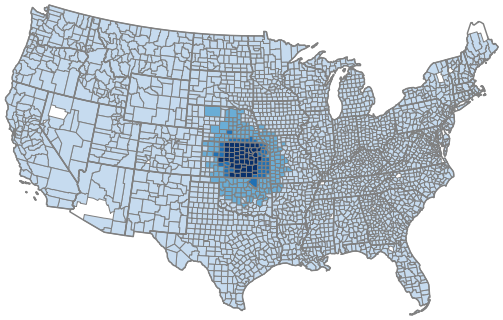
21.A)



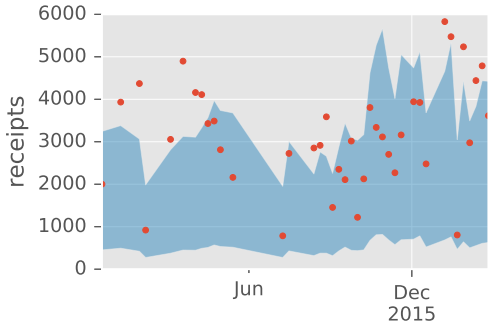
21.B)



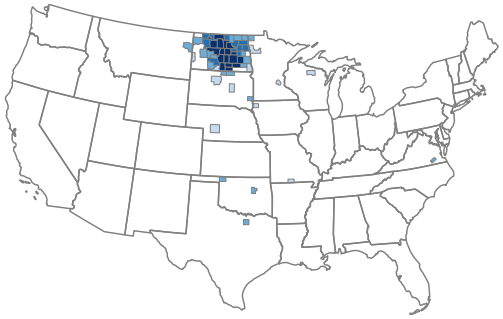
21.C)



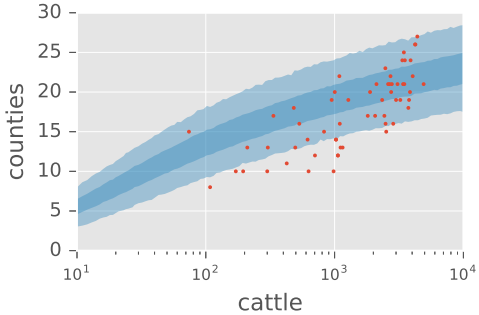
21.D)



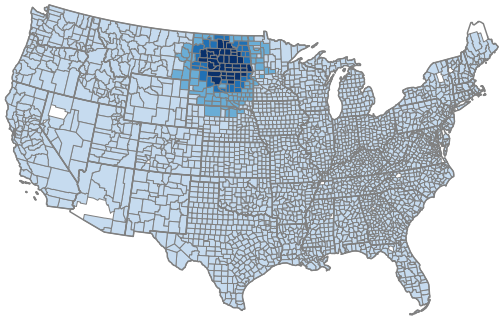
22.A)



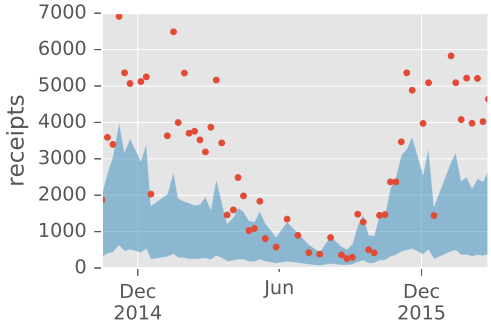
22.B)



22.C)

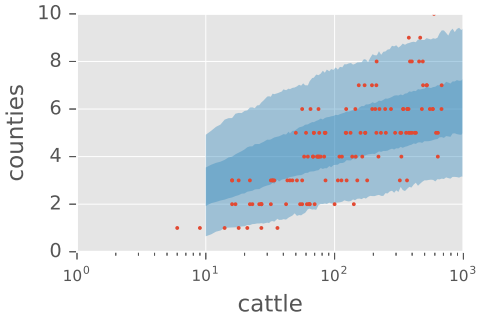


22.D)

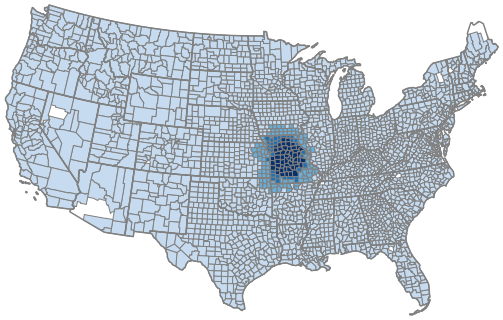


23.A)

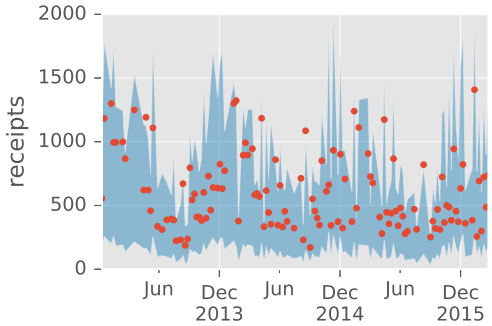
23.B)



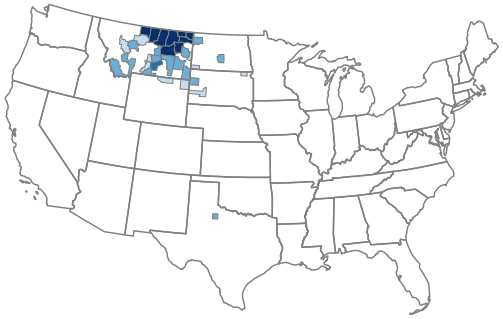
23.C)



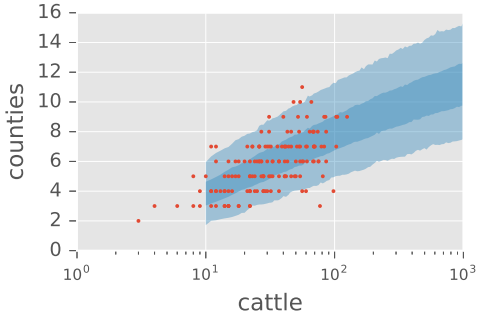
23.D)



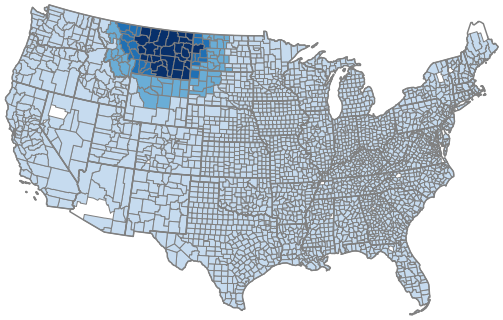
24.A)



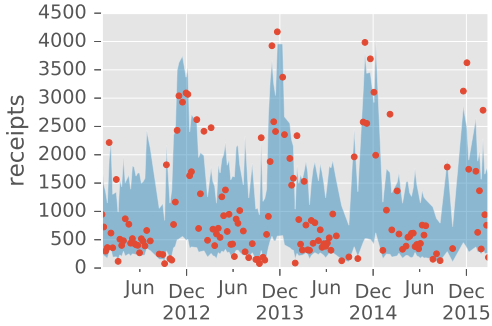
24.B)



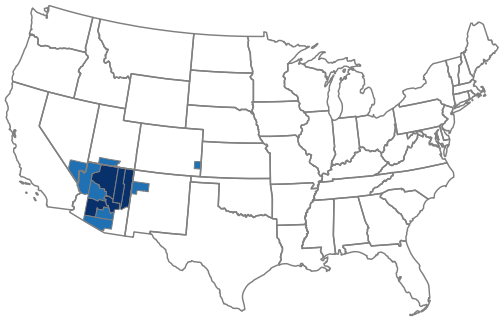
24.C)



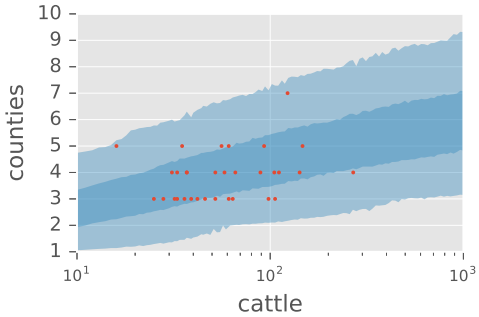
24.D)



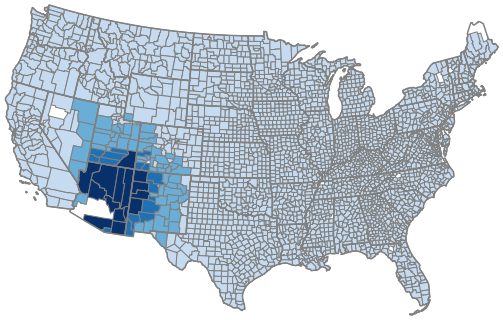
25.A)



25.B)

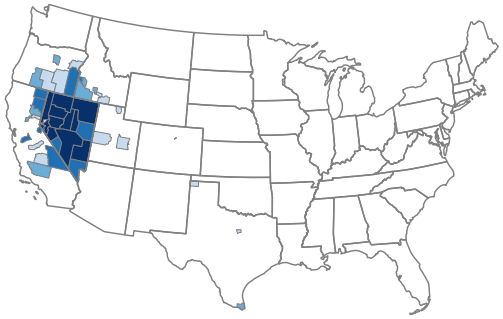


25.C)

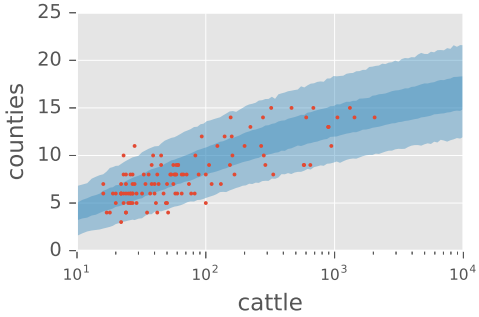


25.D)

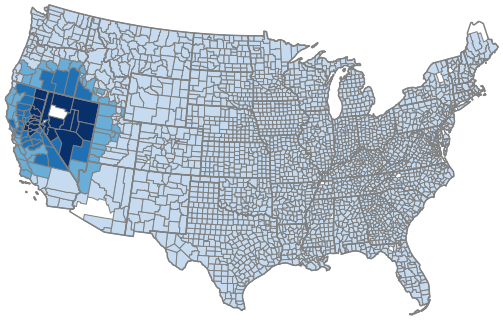
26.A)



26.B)



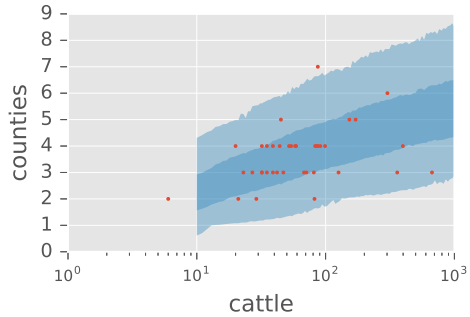
26.C)



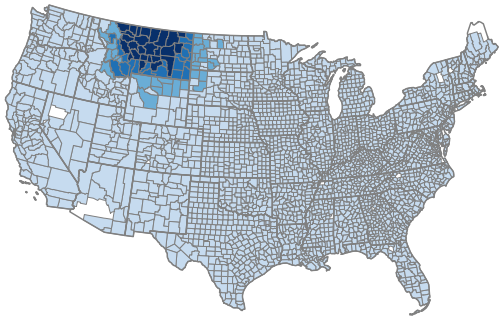
26.D)

27.A)

27.B)

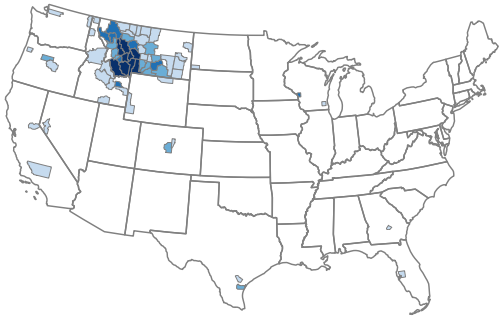


27.C)

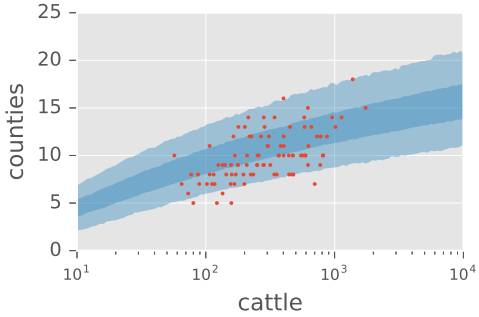


27.D)

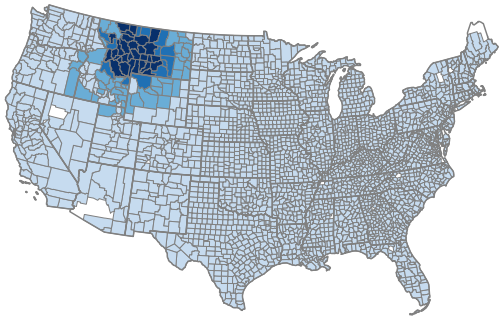
28.A)



28.B)

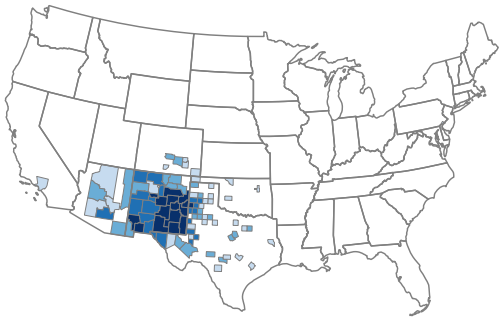


28.C)

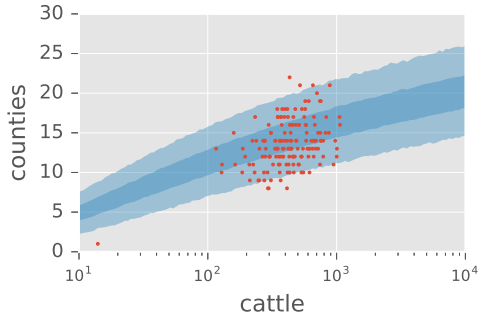


28.D)

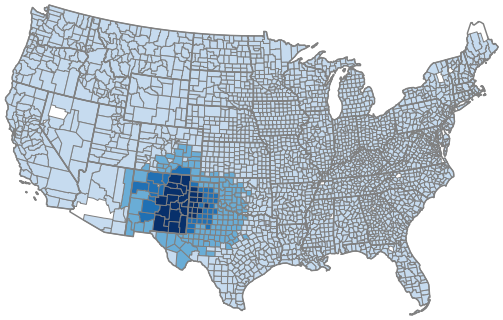
29.A)



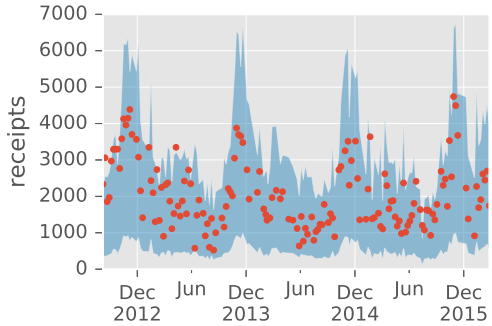
29.B)



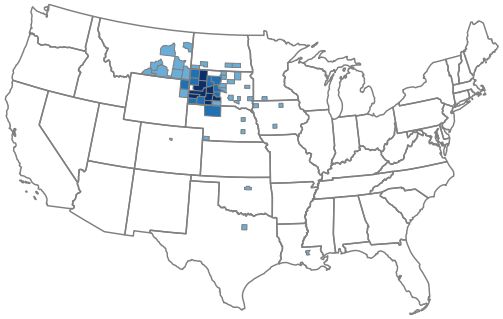
29.C)



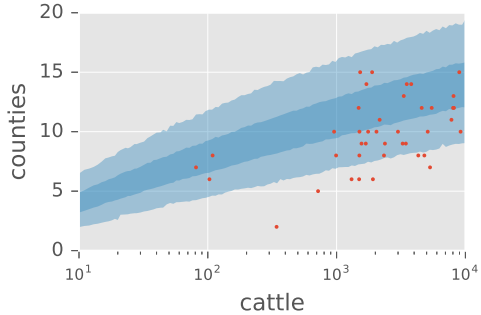
29.D)



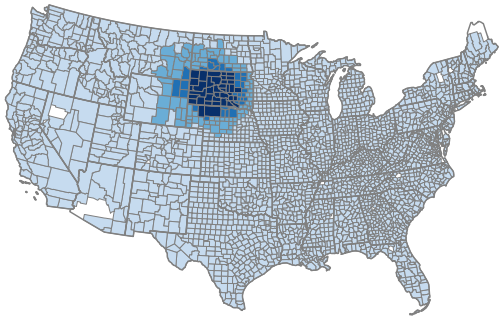
30.A)



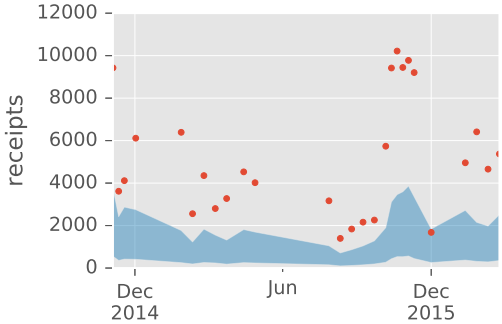
30.B)



30.C)

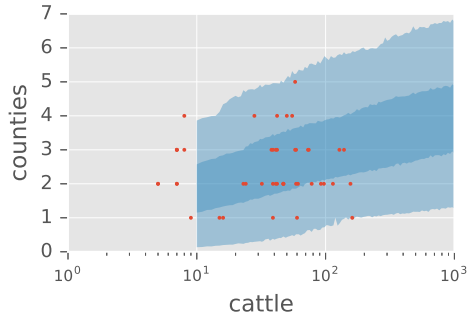


30.D)

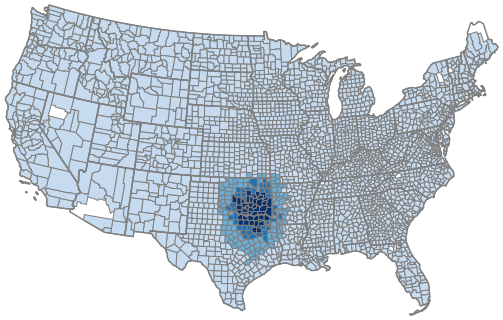


31.A)

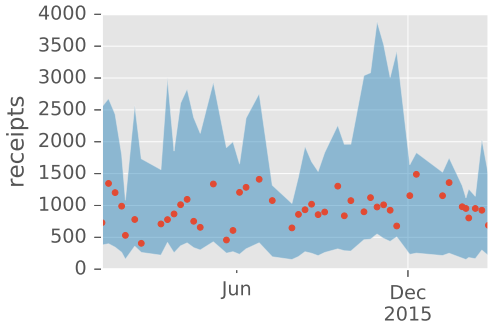
31.B)



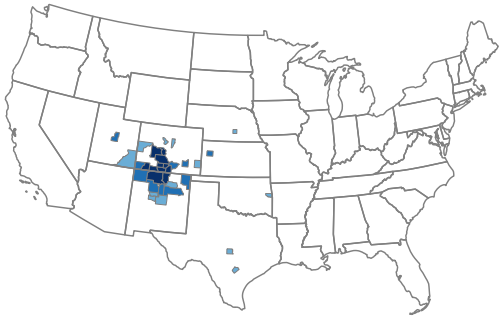
31.C)



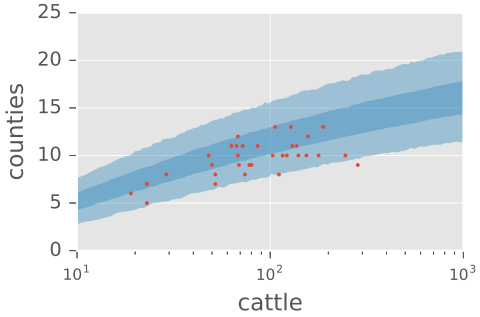
31.D)



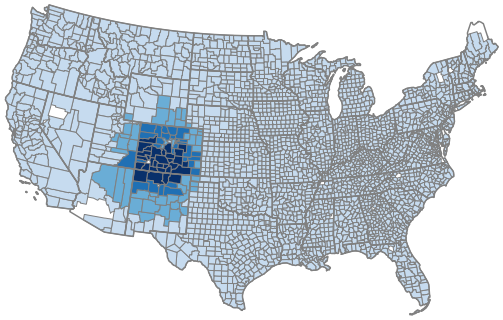
32.A)



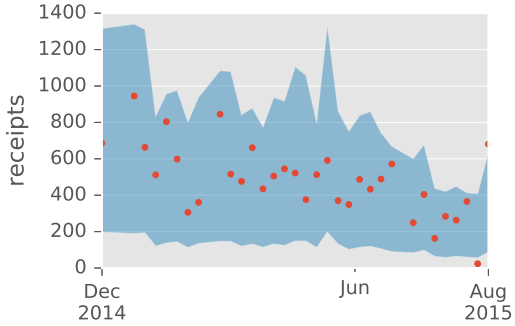
32.B)



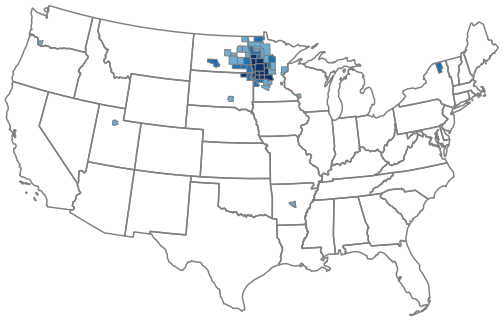
32.C)



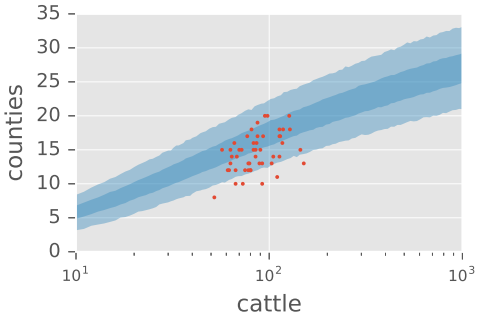
32.D)



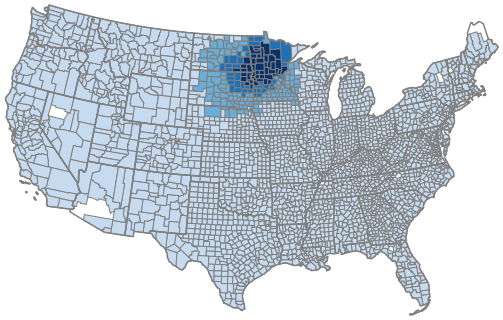
33.A)



33.B)



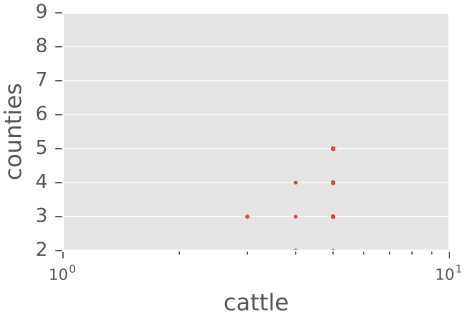
33.C)



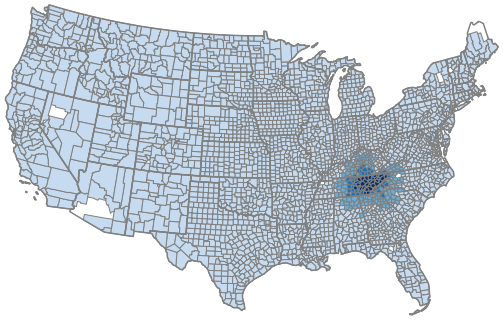
33.D)

34.A)

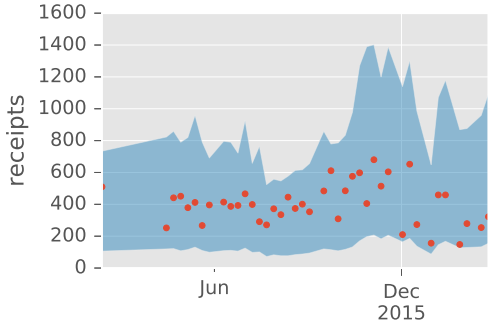
34.B)



34.C)

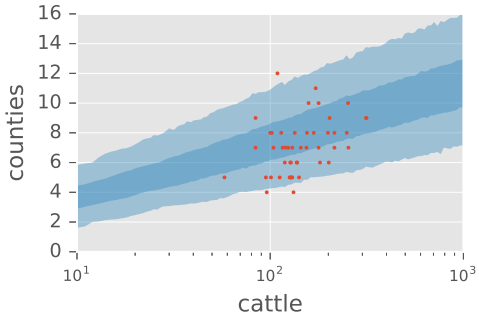


34.D)

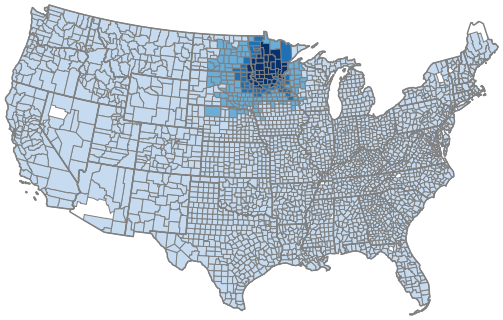


35.A)

35.B)

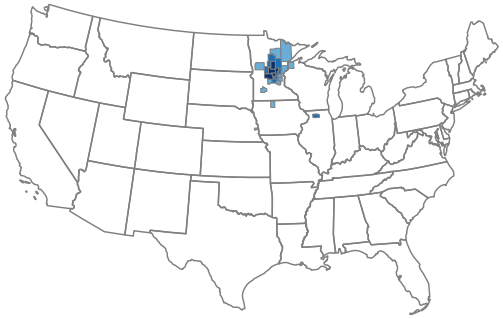


35.C)

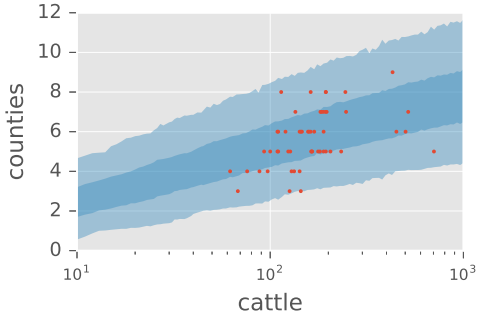


35.D)

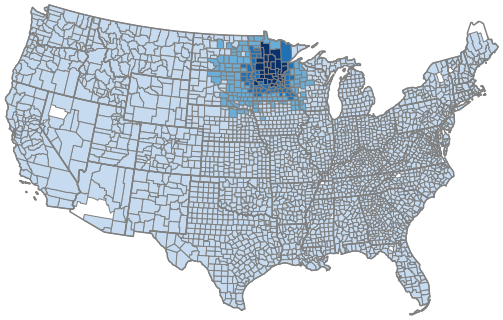
36.A)



36.B)

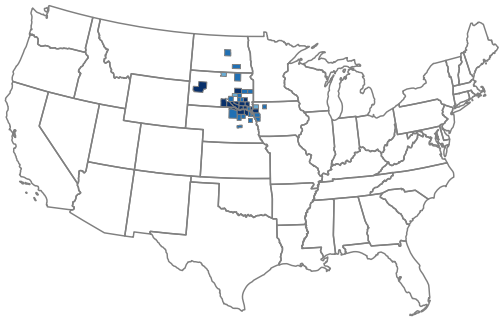


36.C)

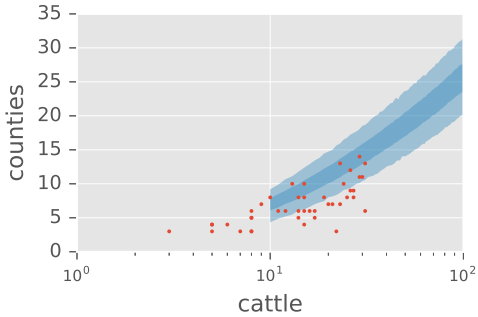


36.D)

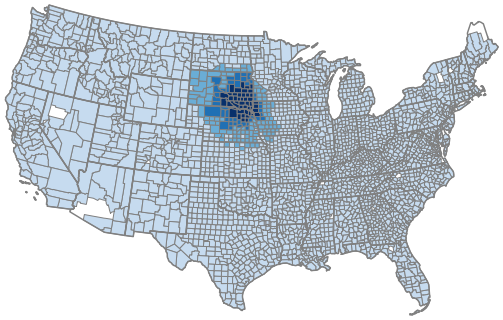
37.A)



37.B)

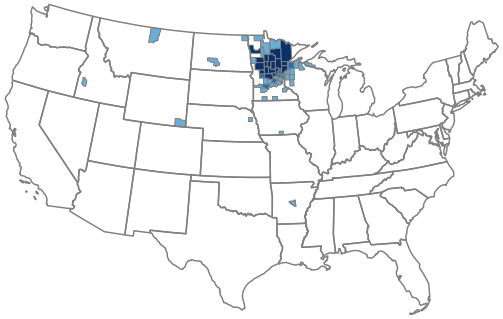


37.C)

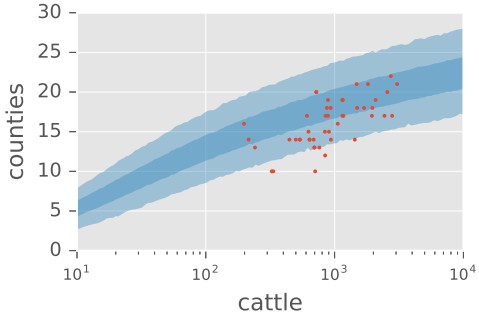


37.D)

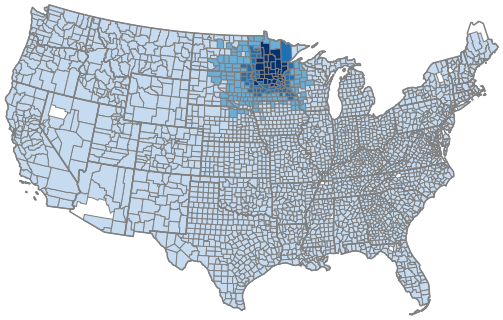
38.A)



38.B)

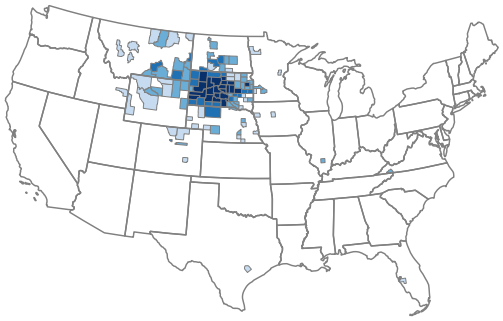


38.C)

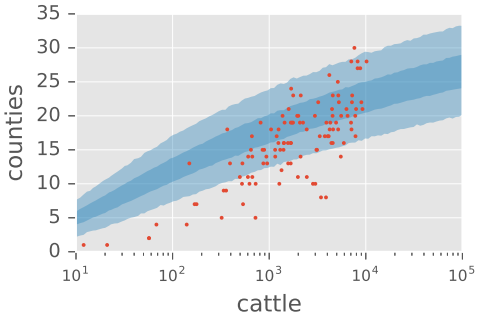


38.D)

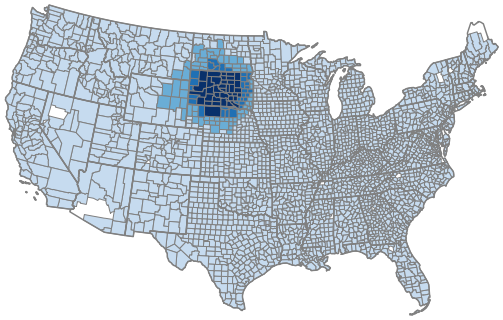
39.A)



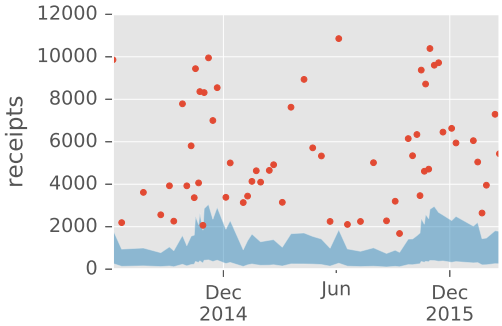
39.B)



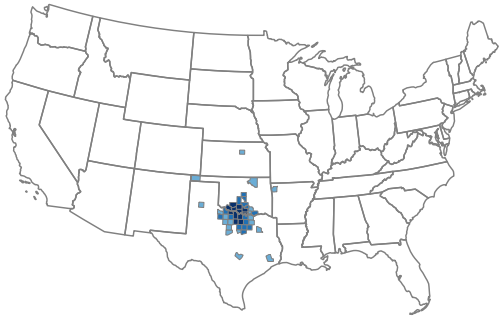
39.C)



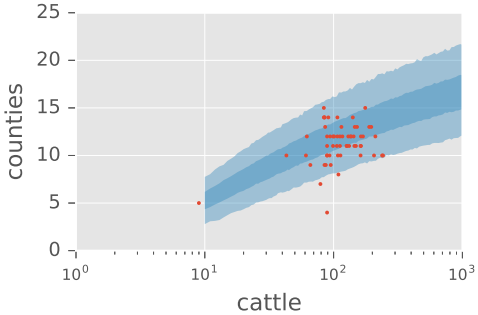
39.D)



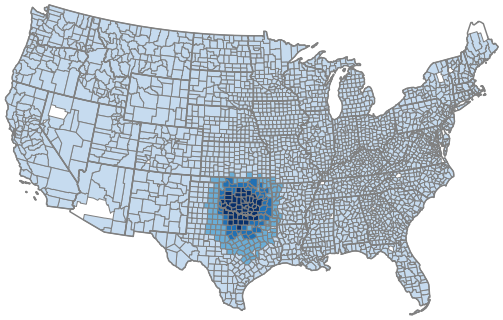
40.A)



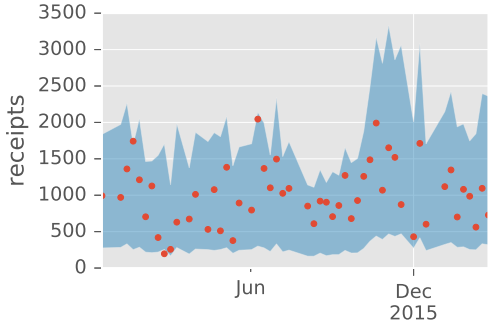
40.B)



40.C)

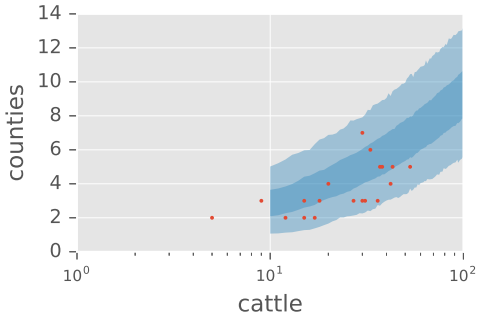


40.D)

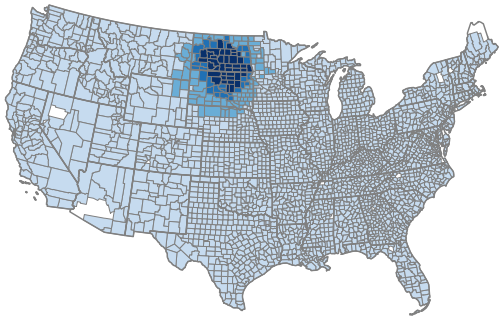


41.A)

41.B)

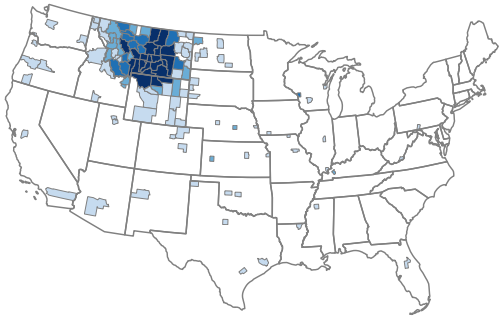


41.C)

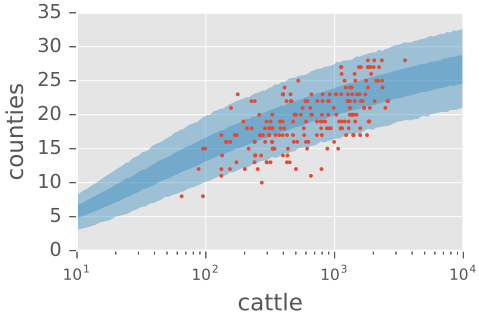


41.D)

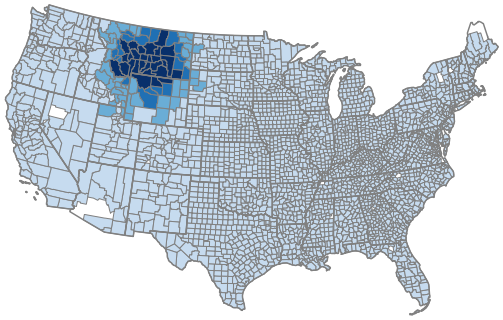
42.A)



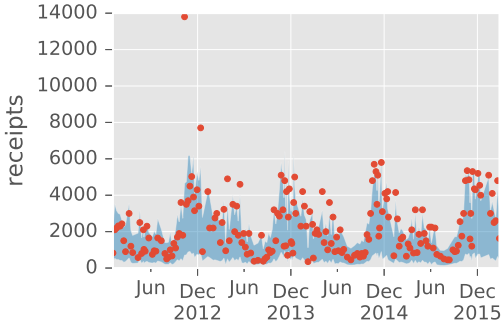
42.B)



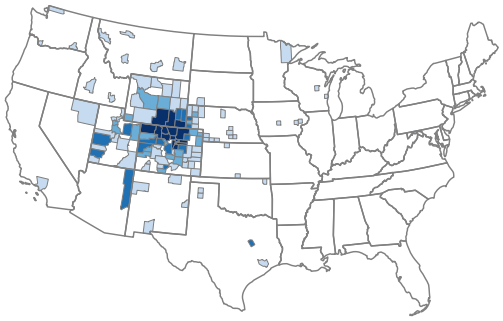
42.C)



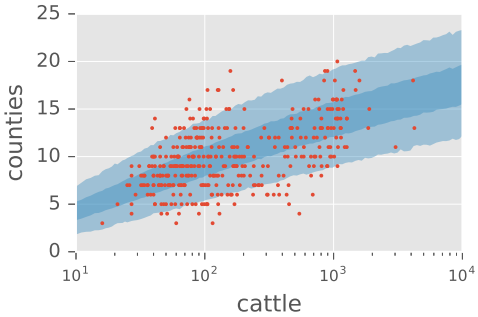
42.D)



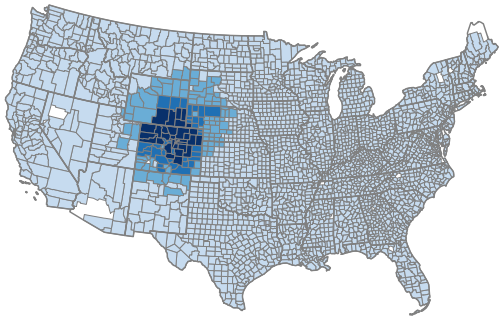
43.A)



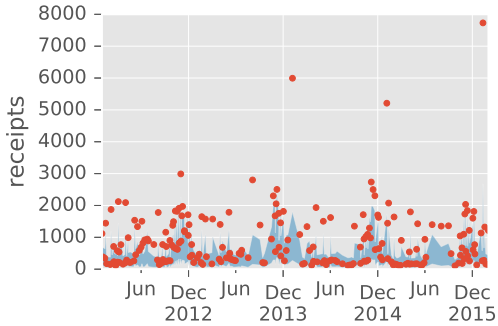
43.B)



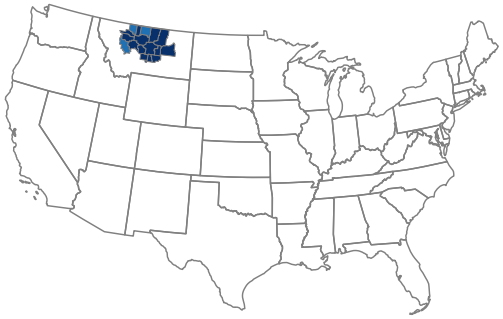
43.C)



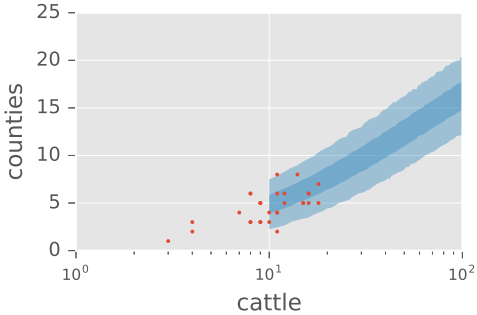
43.D)



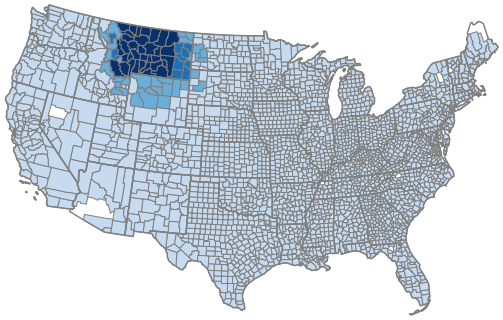
44.A)



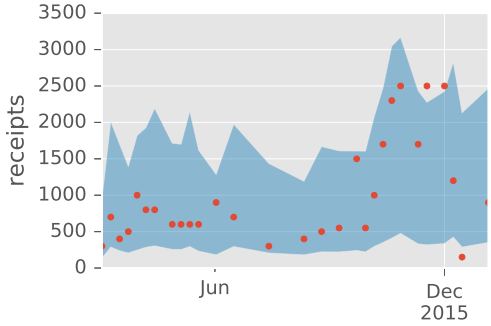
44.B)



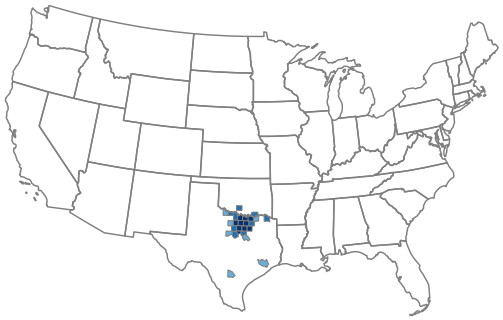
44.C)



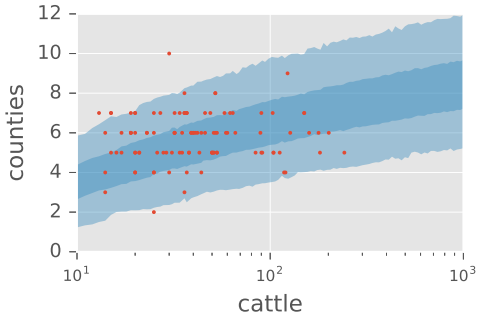
44.D)



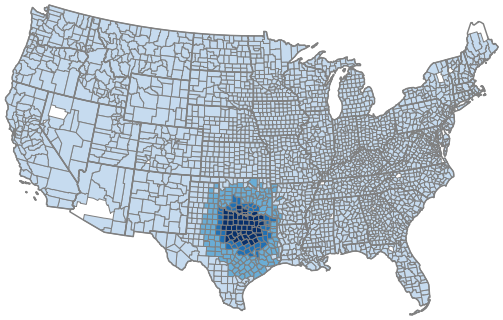
45.A)



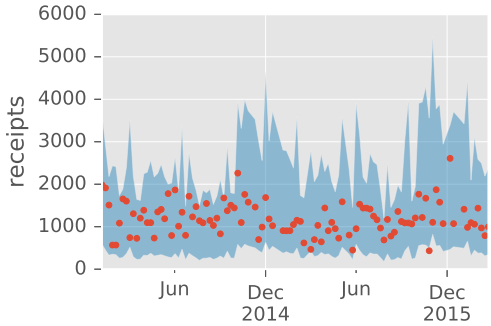
45.B)



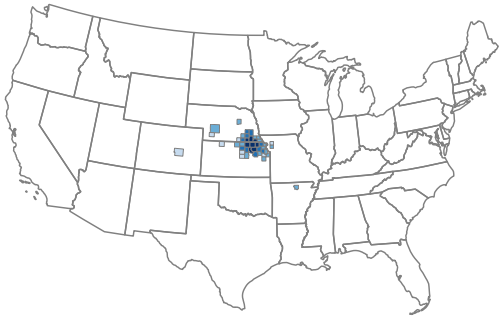
45.C)



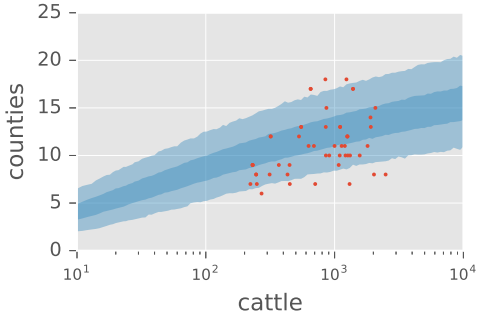
45.D)



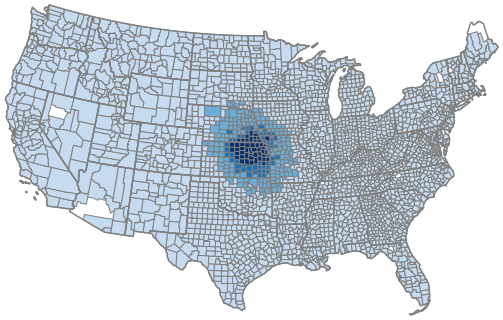
46.A)



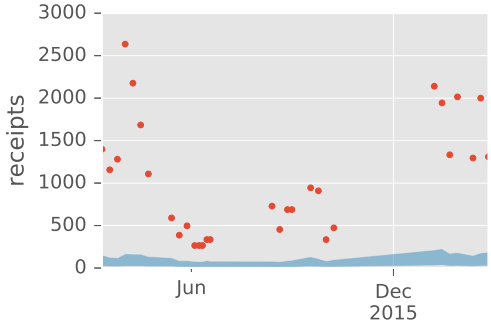
46.B)



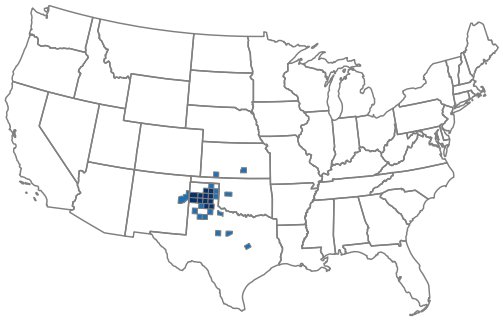
46.C)



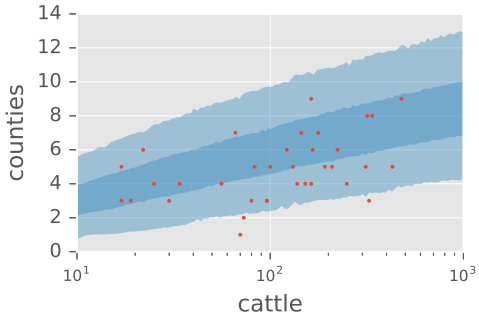
46.D)



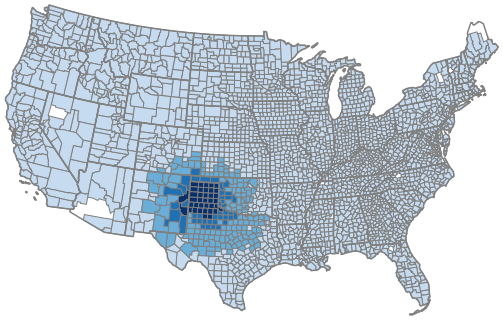
47.A)



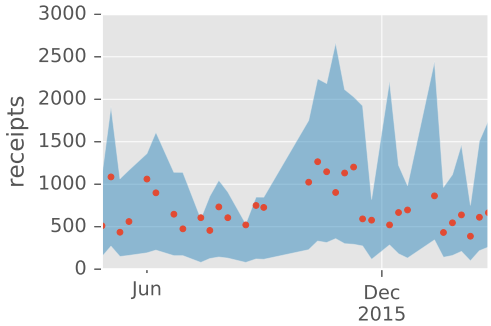
47.B)



47.C)

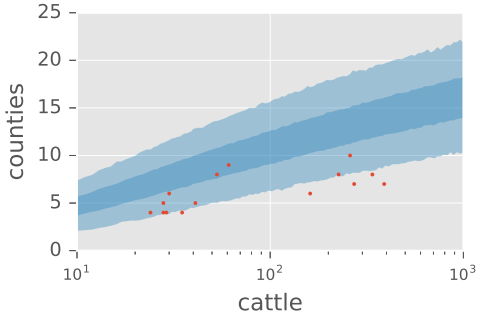


47.D)

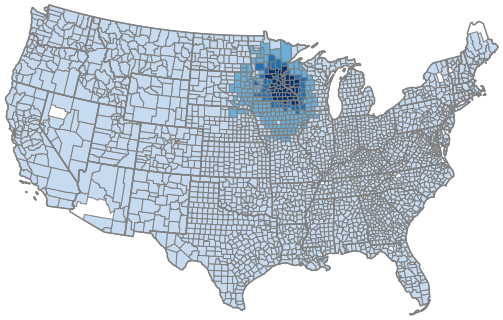


48.A)

48.B)

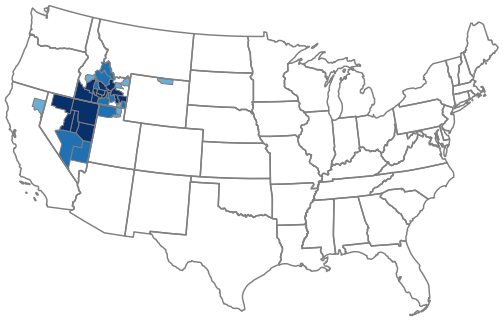


48.C)

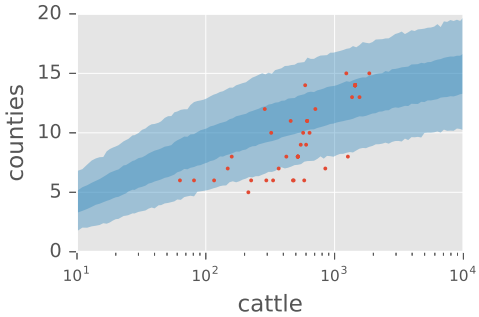


48.D)

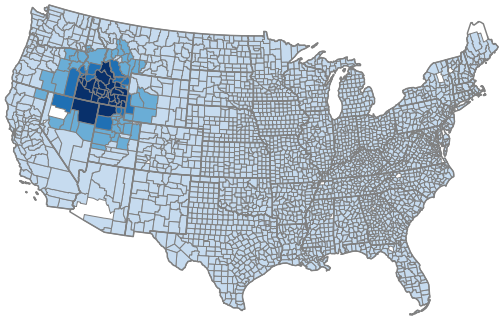
49.A)



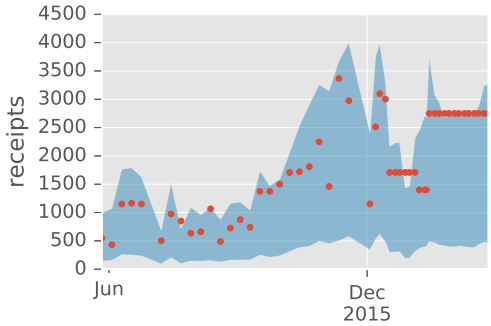
49.B)



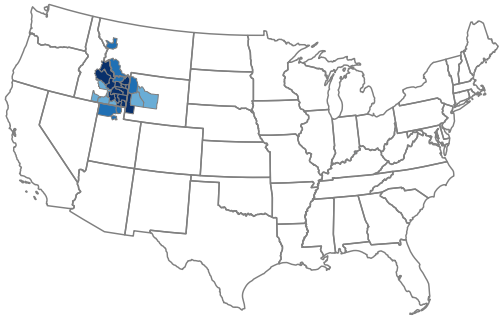
49.C)



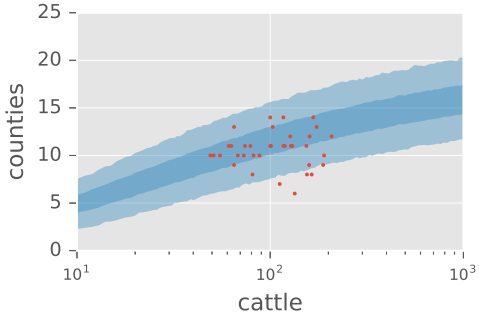
49.D)



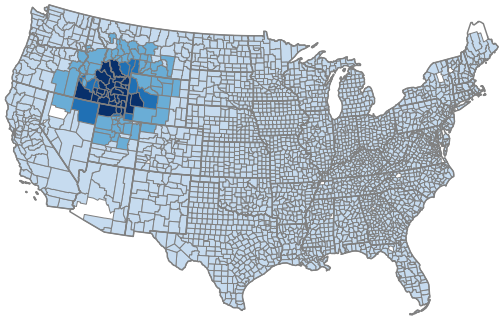
50.A)



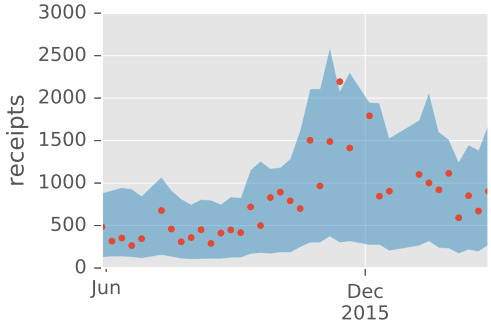
50.B)



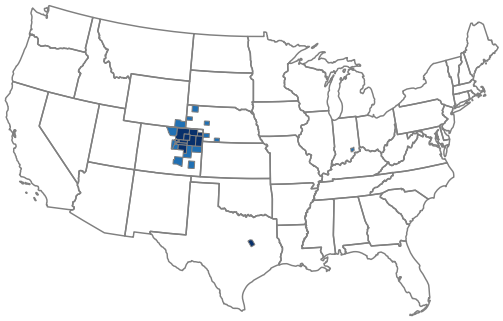
50.C)



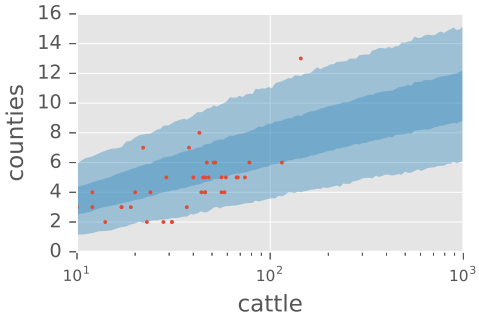
50.D)



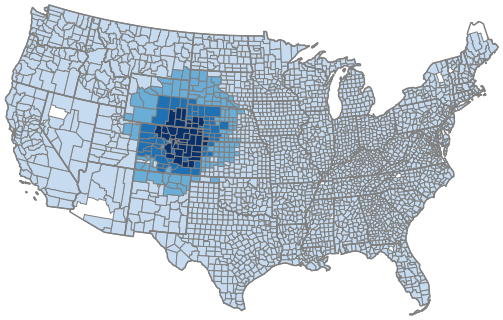
51.A)



51.B)

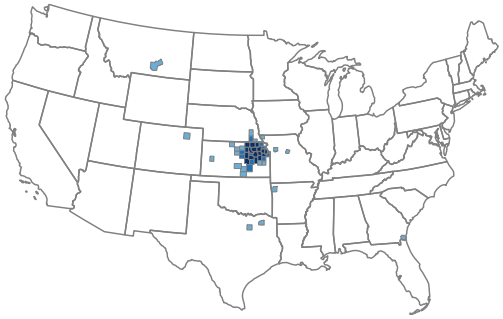


51.C)

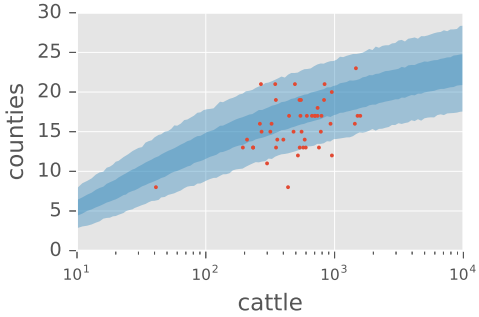


51.D)

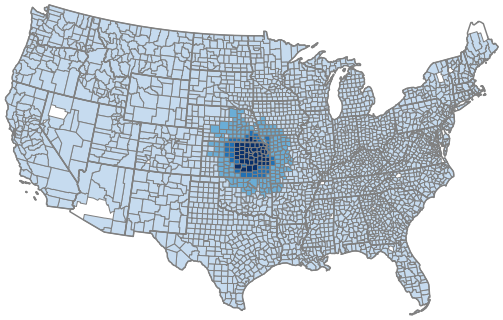
52.A)



52.B)

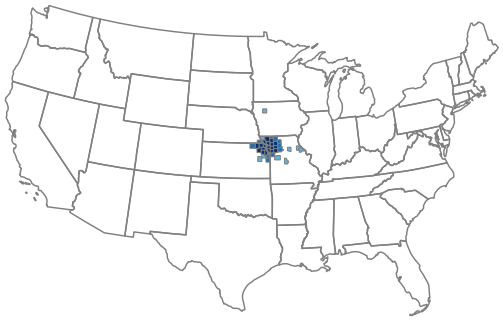


52.C)

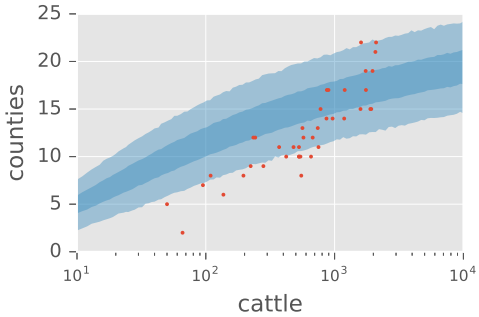


52.D)

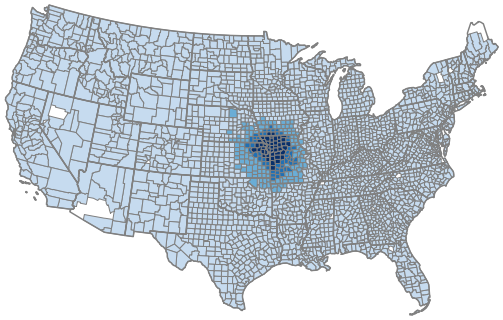
53.A)



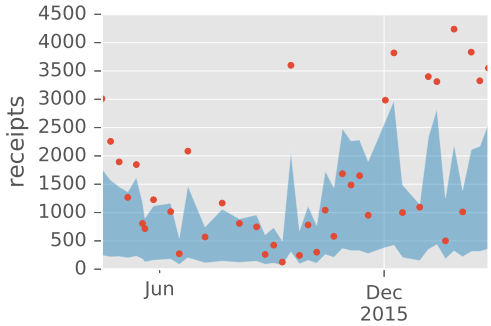
53.B)



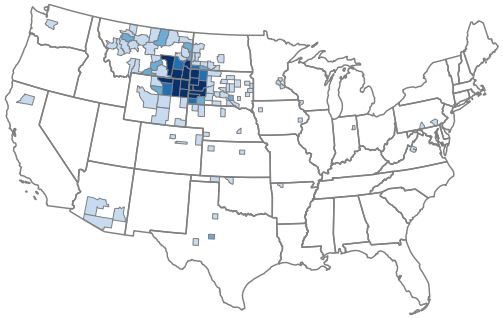
53.C)



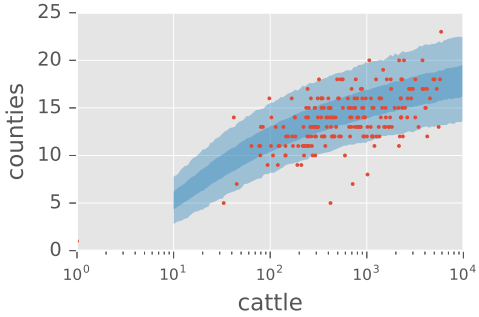
53.D)



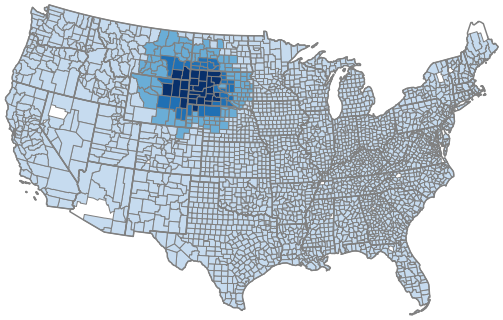
54.A)



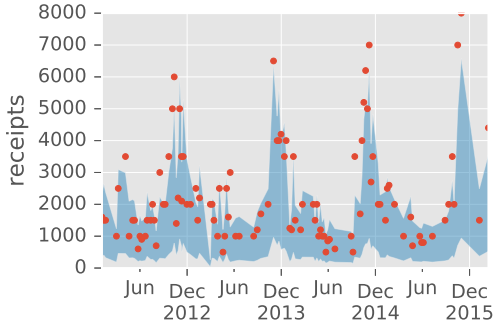
54.B)



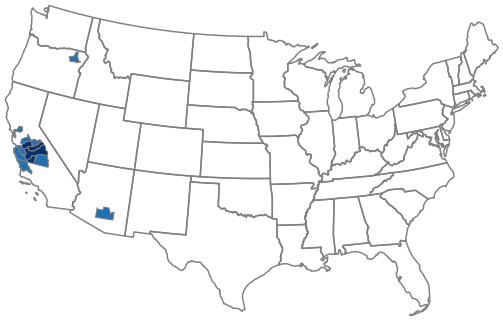
54.C)



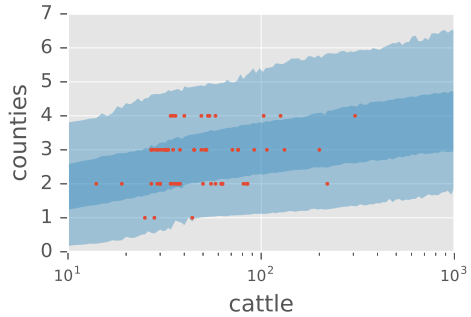
54.D)



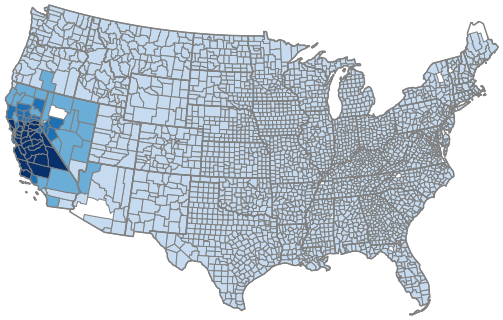
55.A)



55.B)



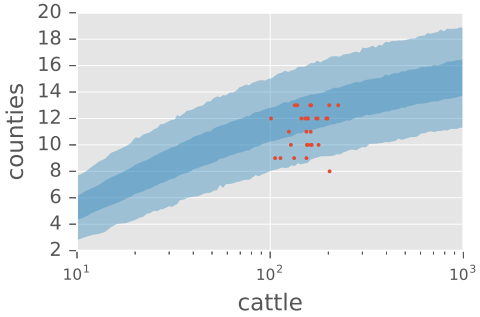
55.C)



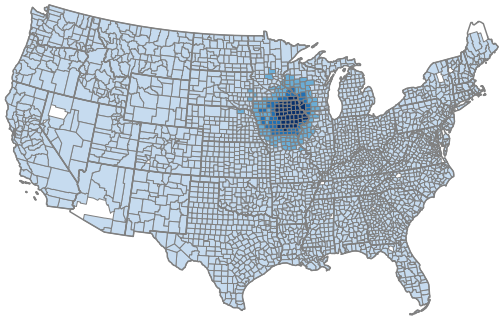
55.D)

56.A)

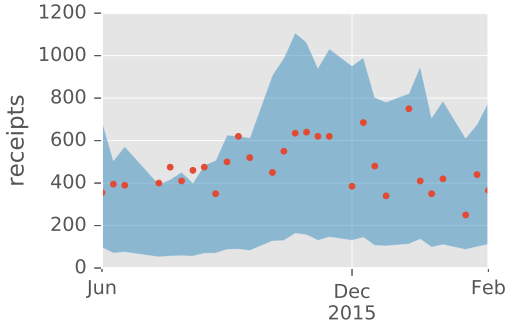
56.B)



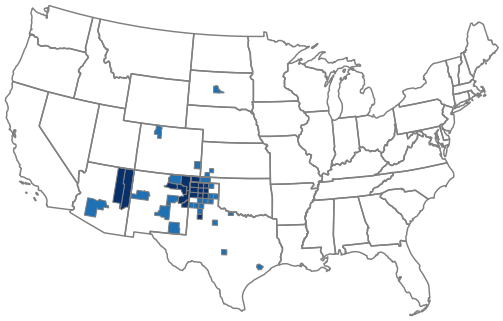
56.C)



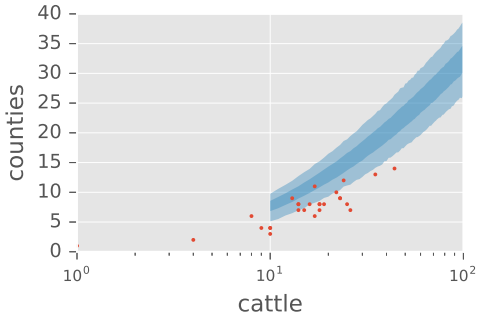
56.D)



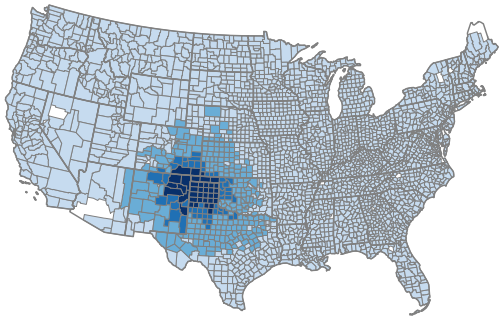
57.A)



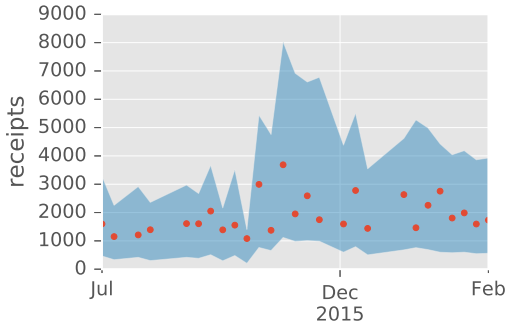
57.B)



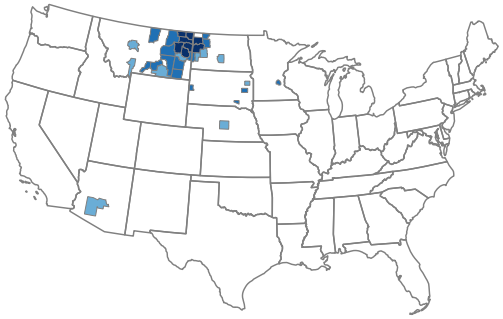
57.C)



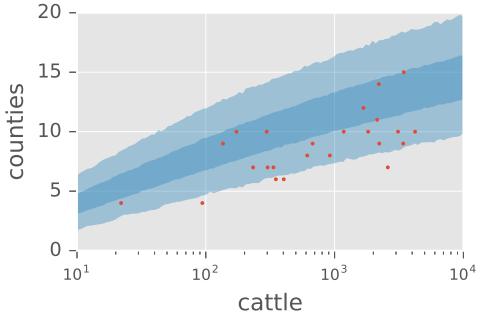
57.D)



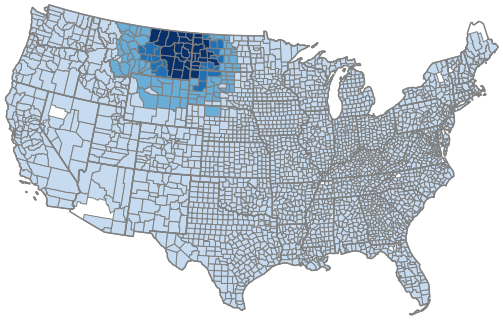
58.A)



58.B)

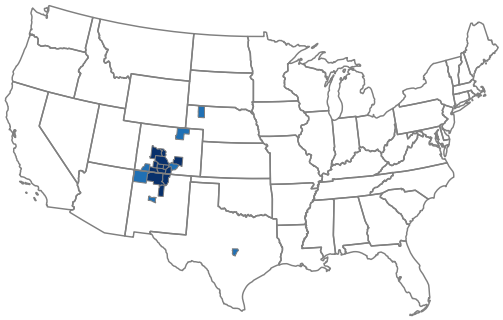


58.C)

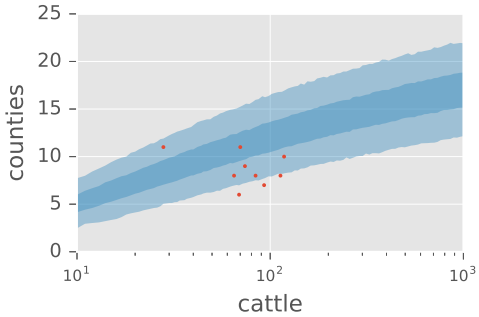


58.D)

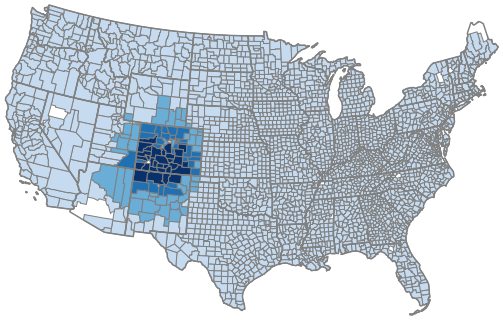
59.A)



59.B)



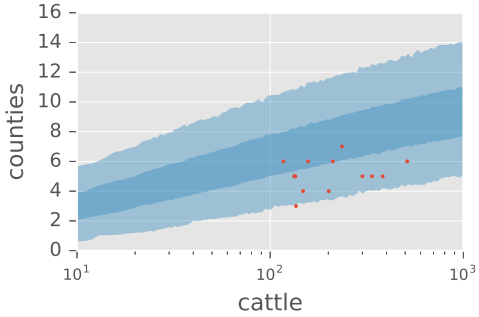
59.C)



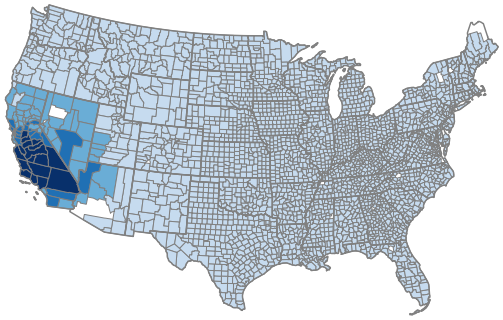
59.D)

60.A)

60.B)

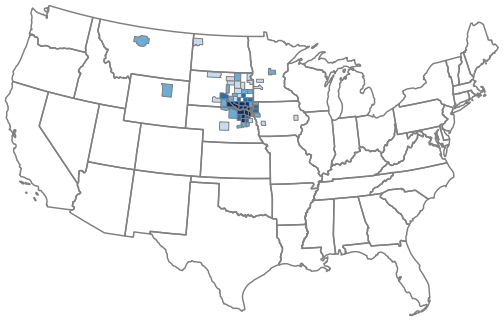


60.C)

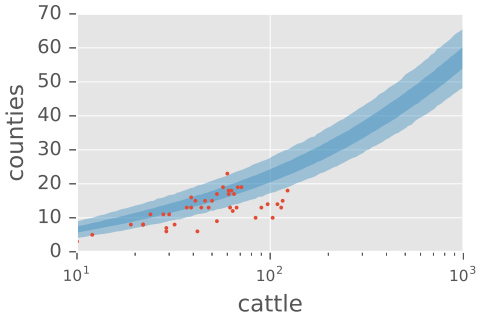


60.D)

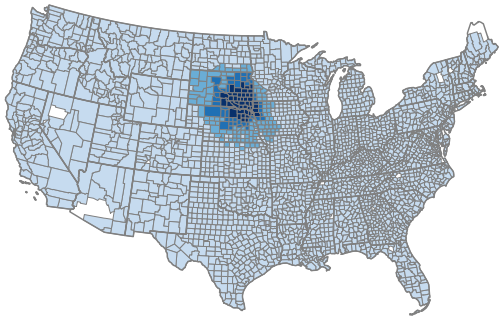
61.A)



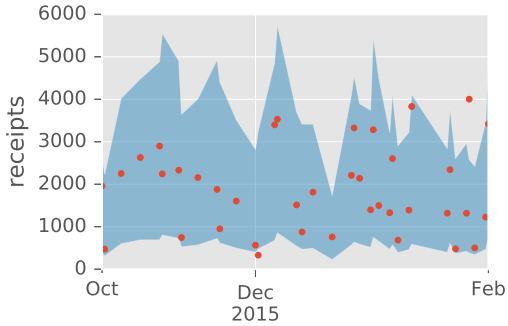
61.B)



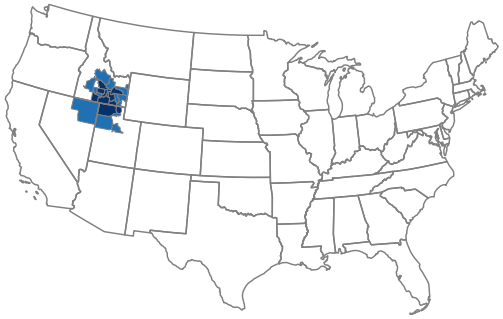
61.C)



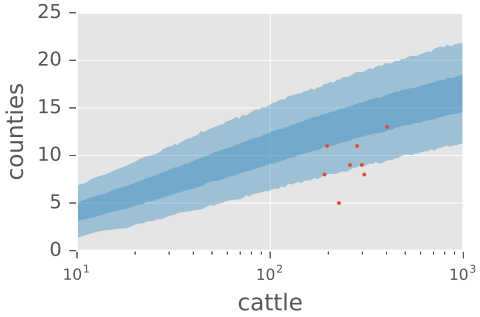
61.D)



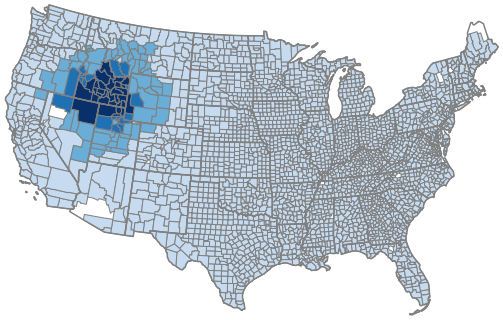
62.A)



62.B)

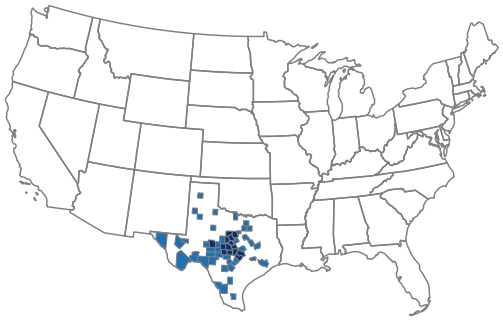


62.C)

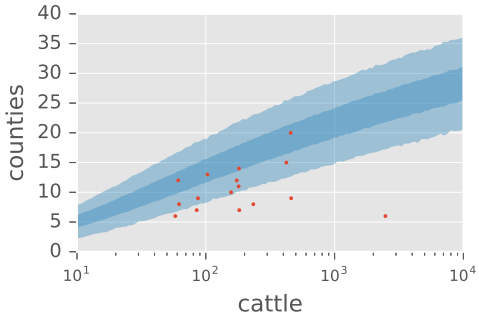


62.D)

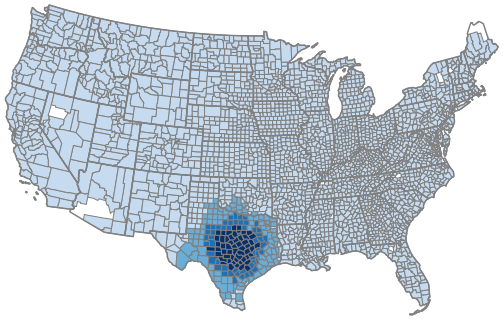
63.A)



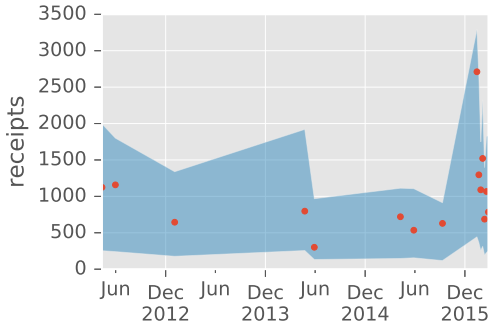
63.B)



63.C)

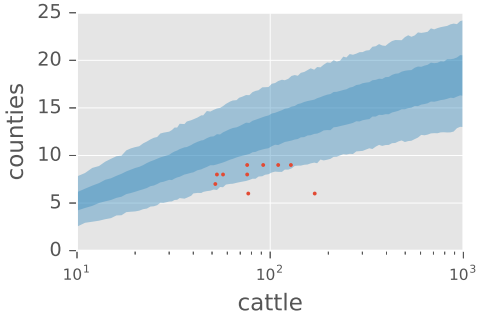


63.D)

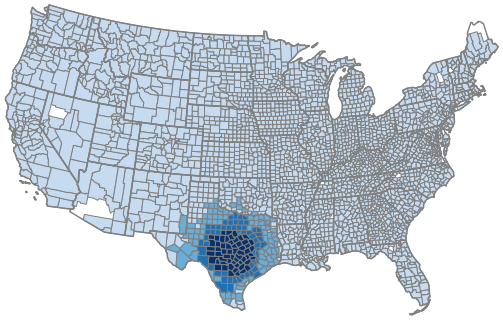


64.A)

64.B)



64.C)



64.D)

Figure C.2: For each market, panels A, B and D correspond to analogous figures for a Montana market shown in the main text (Fig. 2A, Fig. 2B, and Fig. 4A, respectively). Panel C shows the posterior mean values of $e^{\boldsymbol{x}_{i,j}\boldsymbol{\beta}_x}$, binned according to the same boundaries separating quartiles in panel A. Even negligible values in Panel D fall into the first quartile, whereas estimates for $\alpha_{i,j}$ in panel A do not exist if county j was never observed to consign cattle at market i. Missing panels are due to receipts being unavailable for a market.

References

[1] Mathews KH, Brorsen W, Hahn WF, Arnade C, Dohlman E. Mandatory Price Reporting , Market Efficiency , and Price Discovery in Livestock Markets Mandatory Price Reporting , Market Efficiency , and Price Discovery in Livestock Markets Introduction. 2015;.