Supporting Information for "Urban Water Conservation Policies in the United States"

Jonathan M. Gilligan[1,2], Christopher A. Wold[3], Scott C. Worland[2], John J. Nay[2], David J. Hess[4], George M. Hornberger[2]

[1]Department of Earth & Environmental Sciences, Vanderbilt University, Nashville, Tennessee, USA

[2]Department of Civil & Environmental Engineering, Vanderbilt University, Nashville, Tennessee, USA

[3]Vanderbilt Institute for Energy and Environment, Vanderbilt University, Nashville, Tennessee, USA

[4]Department of Sociology, Vanderbilt University, Nashville, Tennessee, USA

Contents

1. Text S1
2. Figures S1 to S5
3. Tables S1 to S7

Additional Supporting Information (Files uploaded separately)

1. Captions for Datasets S1 to S2
2. Data Analysis Scripts S1 (zip file)

Introduction

This supporting information document presents additional details of the data and analysis.

SI Text

Data

We used VWCI data for 195 cities in 49 states, as shown in Fig. S1 and Table S1.

At the MSA level (Dataset S1, Table S1), our regression analysis used the following six covariates: ln(population), population growth rate between 2010 and 2014, the Köppen aridity index, the fraction of the municipal water supply coming from surface water (hence-

Corresponding author: Jonathan M. Gilligan, jonathan.gilligan@vanderbilt.edu

forth, surface-water fraction), the Cook Partisan Voting Index (PVI), and the per-capita real personal income (RPI) for 2014 normalized for inflation and regional variations in the cost of living. We used the natural logarithm of the population rather than the raw population because the raw population was sharply peaked near 500,000 (Fig. 2).

At the state level (Dataset S2, Table S2), our analysis used the following four covariates: PVI, RPI, the Köppen aridity index, and the surface-water fraction.

Analysis

Diagnostics

Pairwise correlation plots of the posterior probability distributions of regressions parameters (Figures S3–S5) are smooth and show little correlation. The Hamiltonian Monte Carlo calculations proceeded without any divergences or exceessive tree depths, and the Gelman-Rubin $\hat{R}$ potential scale-reduction factor (Tabs S5–S7) converged to $\geq$ 0.999 for each parameter.

Model Selection

Leave-one-out cross-validation (Tab. S3) and the Widely Applicable Information Criterion (Tab. S4) were used for model selection (overdispersed beta-binomial versus binomial and hierarchical versus single-level regressions).

Results

Results of the analysis are summarized in Tables S5–7.

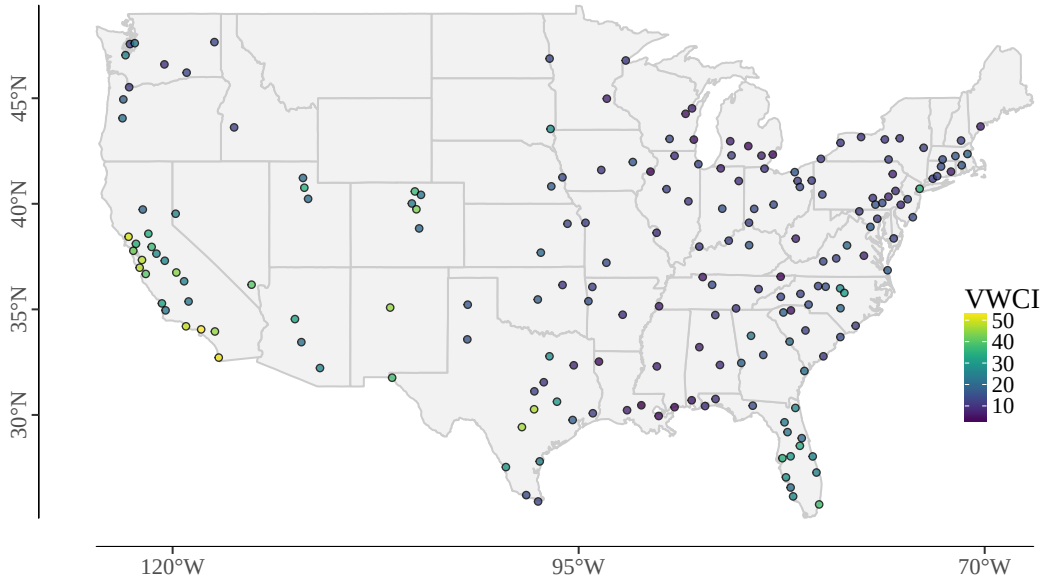

Figure 1. Map of cities with VWCI scores.

Figures S1–S5

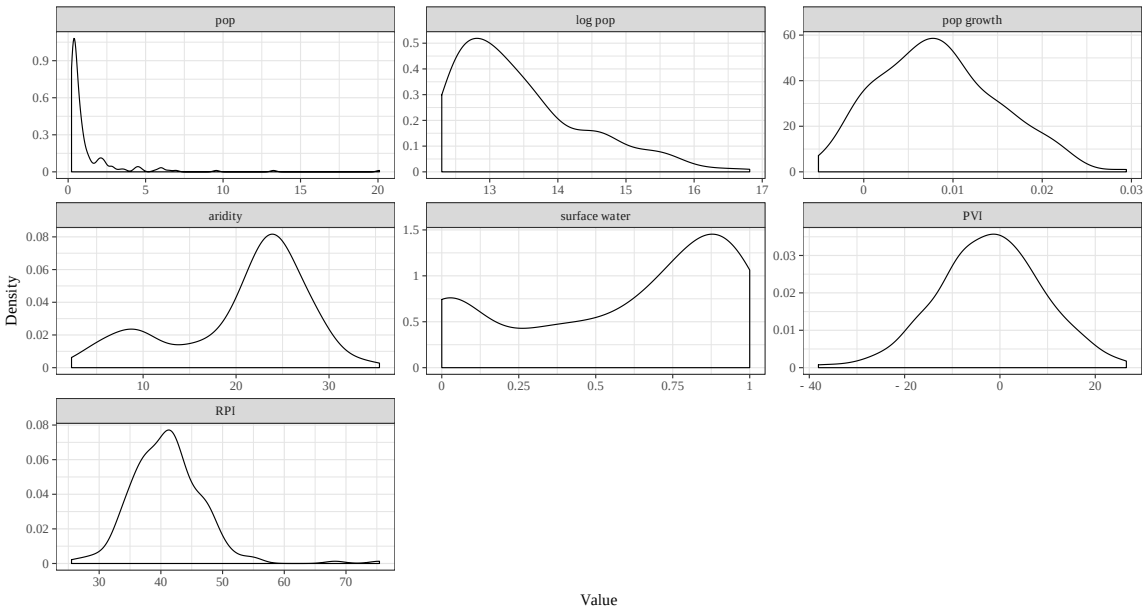

Figure 2. Kernel-density distribution of MSA-level covariates. Population in millions and RPI in thousands of chained 2009 dollars.

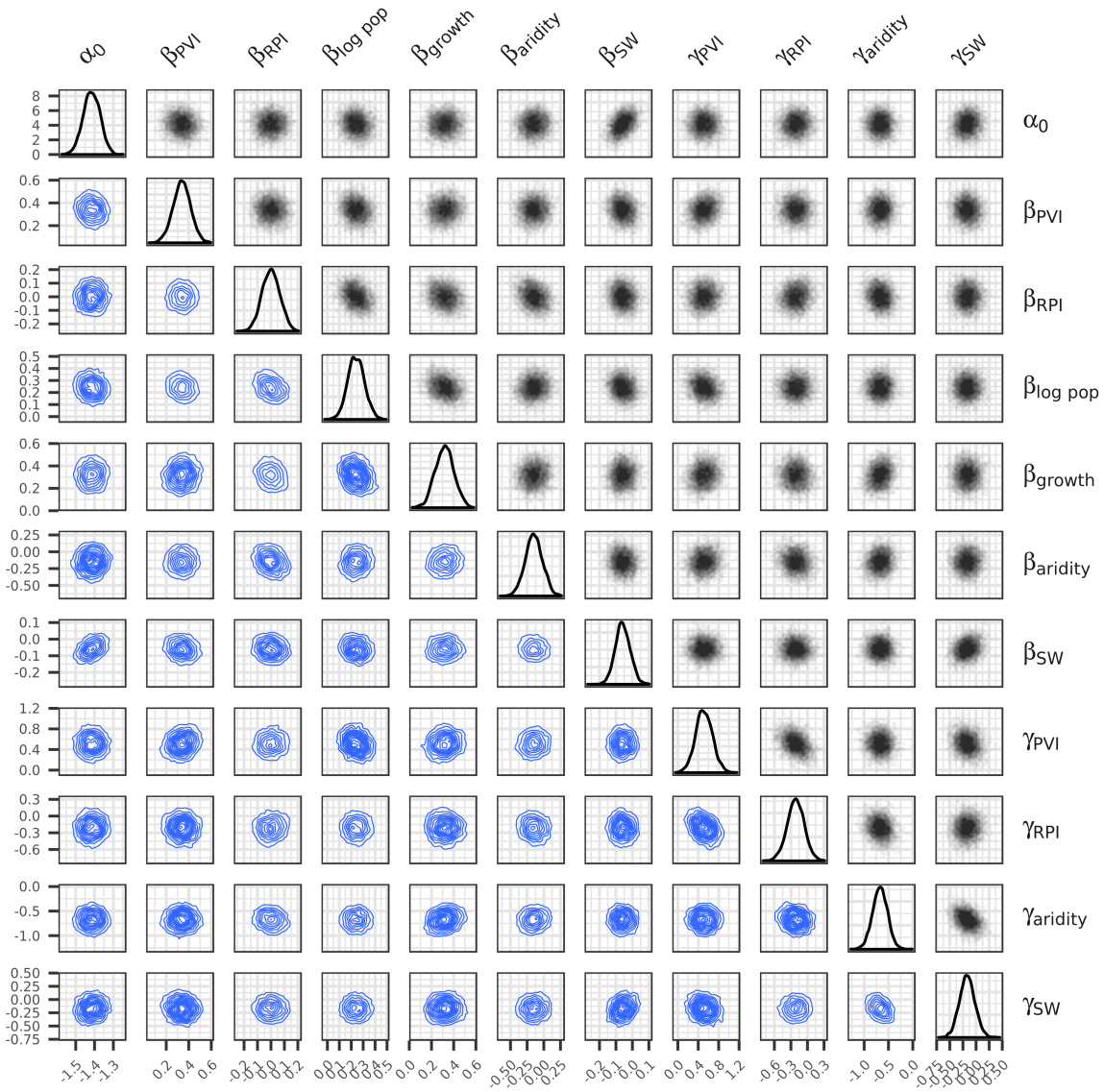

Figure 3. Correlation plot of posterior probability distribution of regression coefficients α, β, and γ for VWCI. The diagonal panels show the probability density for each coefficient, panels in the upper triangle show scatterplots of 4000 HMC samples, and panels in the lower triangle show joint probability density contours corresponding to the scatterplot in the upper triangle. Slight correlations are apparent, as between γ_{aridity} and γ_{SW}, γ_{PVI} and γ_{RPI}, and β_{SW} and α_0, but these are small enough not to pose problems apart from slightly increasing the uncertainty in the parameter estimates.

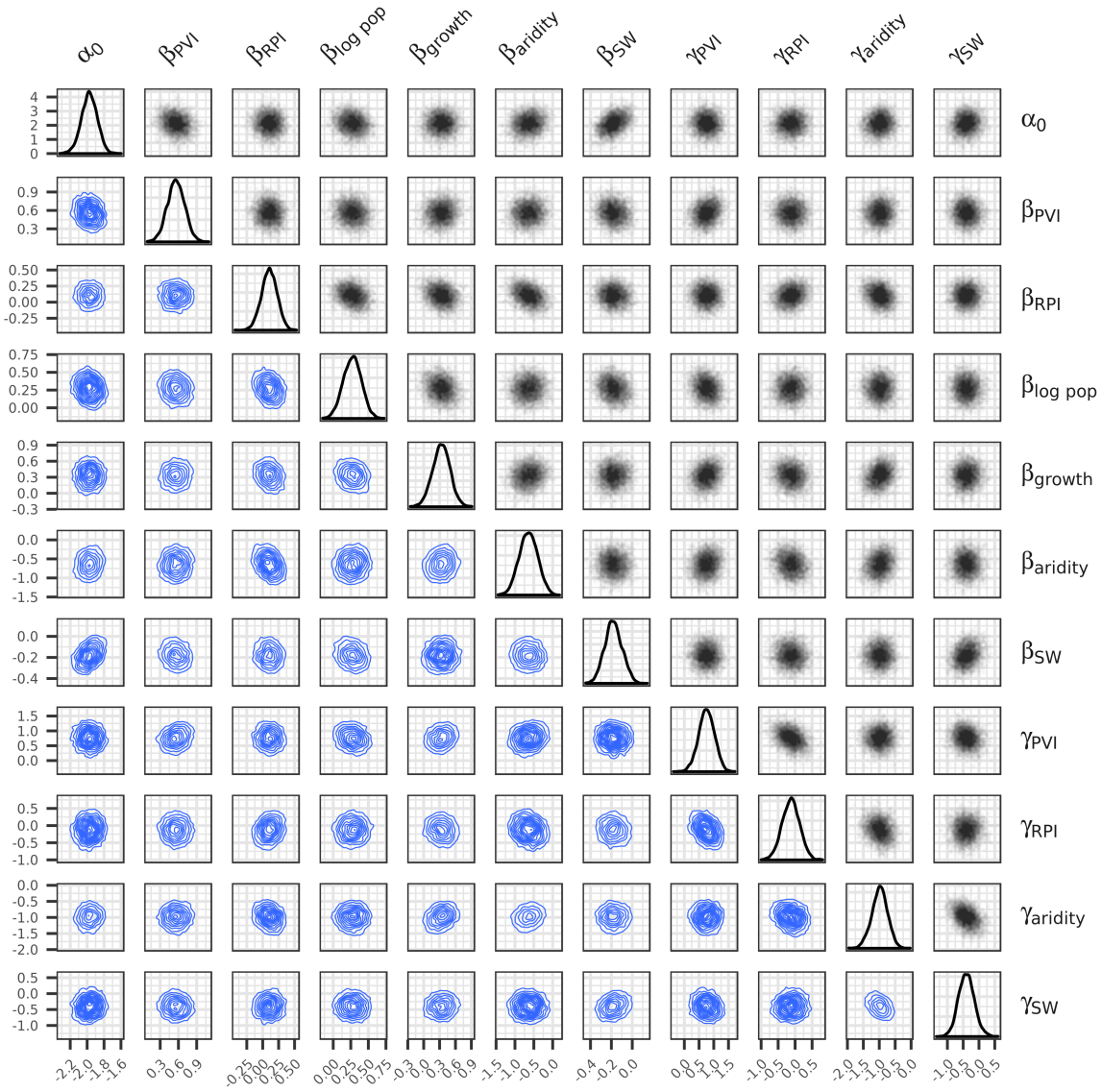

Figure 4. Correlation plot of posterior probability distribution of regression coefficients α, β, and γ for requirements.

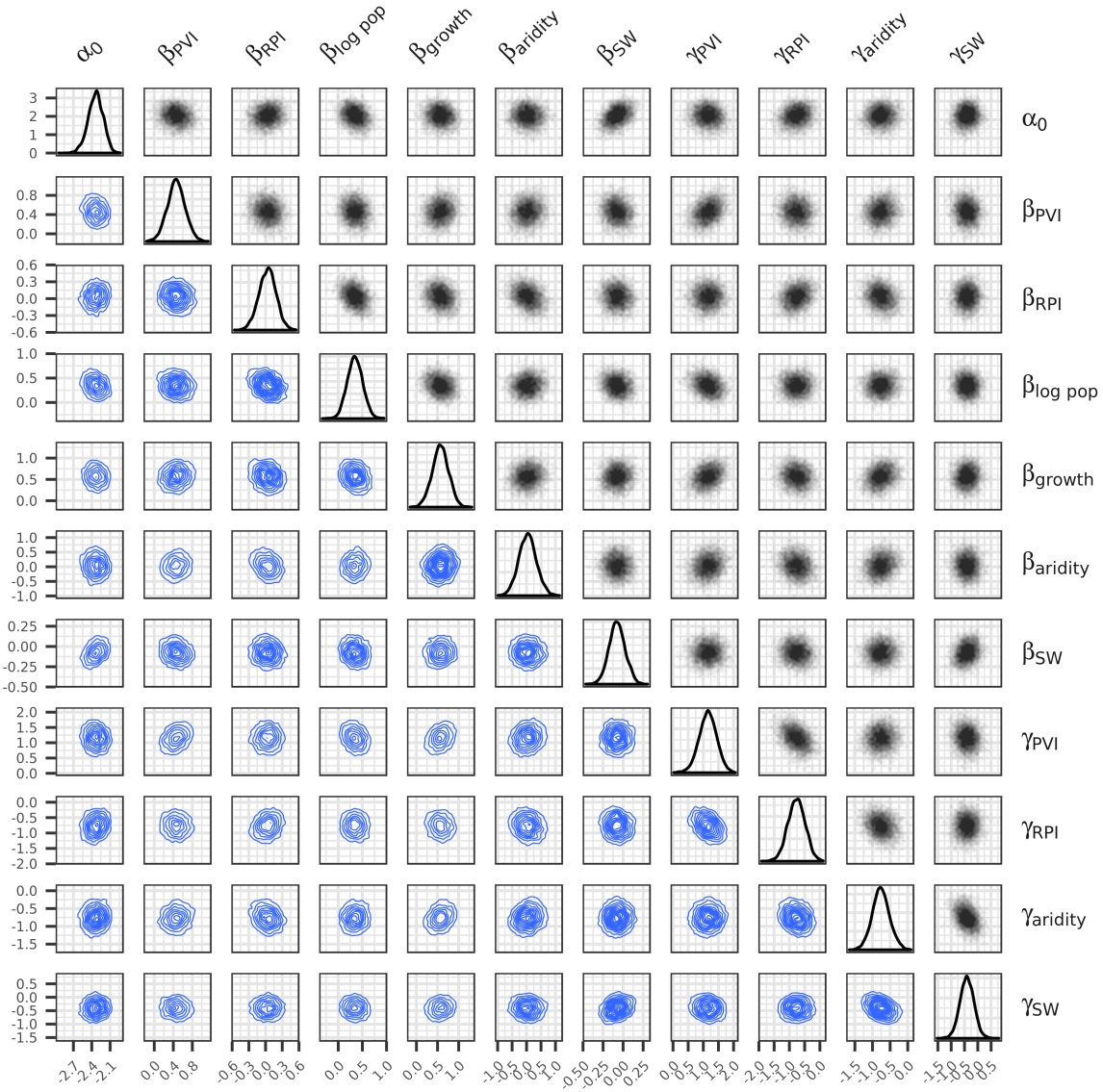

Figure 5. Correlation plot of posterior probability distribution of regression coefficients α, β, and γ for rebates.

1 Tables S1–S7

Table S1 Caption

Table 1. Conservation scores and covariates for cities: VWCI = Vanderbilt Water Conservation Index (total # of conservation measures), Req. = # requirements, Reb. = # rebates, PVI = Cook Partisan Voting Index, Aridity = Köppen aridity index, RPI = per-capita real personal income (thousands of regionally adjusted chained 2009 dollars), Pop. = population (thousands), Growth = population growth rate (2010–2014), Surf. W. = surface-water fraction.

Table S2 Caption

Table 2. State-level covariates: PVI = Cook Partisan Voting Index, RPI = per-capita real personal income (thousands of regionally-adjusted chained 2009 dollars), Aridity = the Köppen aridity index, Surf. W. = the surface-water fraction.

Table S3

Model	LOO-IC	s.e. LOO-IC	$ELPD_{LOO}$	s.e. $ELPD_{LOO}$
hierarchical beta-binomial	1249.4	20.5	-624.7	10.2
hierarchical binomial	1334.3	44.3	-667.2	22.2
single-level binomial	2064.6	143.8	-1032.3	71.9
single-level beta-binomial	2438.6	183.4	-1219.3	91.7

Table 3. Model comparison: LOO = leave-one-out cross-validation, LOO-IC = LOO information criterion, ELPD = expected log pointwise predictive density, and s.e. indicates the standard error of estimates of quantities. Lower values of the information criteria and greater (less negative) values of ELPD indicate superior model performance.

Table S4

Model	WAIC	s.e. WAIC	$ELPD_{WAIC}$	s.e. $ELPD_{WAIC}$
hierarchical beta-binomial	1245.2	20.1	-622.6	10.1
hierarchical binomial	1320.2	43.5	-660.1	21.7
single-level binomial	2063.2	143.6	-1031.6	71.8
single-level beta-binomial	2386.1	177.5	-1193.1	88.8

Table 4. Model comparison: WAIC = widely applicable information criterion (also known as the Watanabe-Aikake Information Criterion), ELPD = expected log-probability density, and s.e. indicates the standard error of estimates of quantities. Lower values of the information criteria and greater (less negative) values of ELPD indicate superior model performance.

Table S5 Caption

Table 5. Posterior probability distributions of regression coefficients for VWCI: mean, standard error of the mean, standard deviation of the posterior, quantiles of the posterior, and the Gelman-Rubin potential scale-reduction factor $\hat{R}$. γ coefficients correspond to state-level effects, β coefficients to MSA-level effects, δ coefficients represent state-level intercepts, α_0 is the overall intercept, and ϕ characterizes the overdispersion of the beta-binomial distribution. For more detail, see Materials and Methods.

Table S6 Caption

Table 6. Posterior probability distribution of regression coefficients for requirements

Table S7 Caption

Table 7. Posterior probability distribution of regression coefficients for rebates

Captions for Datasets S1–S2

Dataset S1: MSA-Level Data

This dataset contains MSA-level data: the FIPS (Federal Information Processing Standard) code for the MSA, the name of the MSA, the central city, state, latitude, longitude, VWCI, number of water-conservation requirements, number of rebate policies for water-conservation actions, the average annual precipitation (in millimeters) temperature (in Celsius), and Köppen aridity index, for the central city, the Cook Partisan Voting Index for the counties of the MSA, the 2014 population and average annual population growth rate from 2010–2014 for the MSA, the fraction of the municipal water supply derived from surface water, the BEA 2014 regional price parity and per-capita real personal income for the MSA (in chained regionally-adjusted 2009 dollars).

Dataset S2: State-Level Data

This dataset contains stae-level data: the FIPS code for the state, the abbreviation and name of the state, the average annual precipitation (in millimeters), temperature (in Celsius), and Köppen aridity index for the state, the state-level Cook Partisan Voting Index, the fraction of the state water supply derived from surface water, and the BEA 2014 state-level regional price parity and per-capita real personal income (in chained regionally-adjusted 2009 dollars).

Data Analysis Scripts S1

The zip file `scripts_S1.zip` contains R and Stan scripts to reproduce the regression analysis presented here. To reproduce the analysis, unzip the file with the scripts, copy Datasets S1 and S2 into the `data` subdirectory, and run the script `gilligan_vwci_ef_2017.R` in R.

This paper was produced with the following R packages:

```
## R version 3.3.3 (2017-03-06)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 16.04.2 LTS
##
## locale:
```

```
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] jgally_1.2.9.9999    jgmcmc_1.1           loo_1.1.0
##  [4] rstan_2.14.2         StanHeaders_2.14.0-1 viridis_0.4.0
##  [7] viridisLite_0.2.0    ggthemes_3.4.0       xtable_1.8-2
## [10] gridExtra_2.2.1      ggrepel_0.6.5        extrafont_0.17
## [13] stringr_1.2.0        readxl_0.1.1         dplyr_0.5.0
## [16] purrr_0.2.2          readr_1.1.0          tidyr_0.6.1
## [19] tibble_1.3.0         ggplot2_2.2.1        tidyverse_1.1.1
## [22] pacman_0.4.1         knitr_1.15.1
##
## loaded via a namespace (and not attached):
##  [1] reshape2_1.4.2     haven_1.0.0        lattice_0.20-34
##  [4] colorspace_1.3-2   stats4_3.3.3       foreign_0.8-67
##  [7] DBI_0.6-1          RColorBrewer_1.1-2 modelr_0.1.0
## [10] matrixStats_0.51.0 plyr_1.8.4         munsell_0.4.3
## [13] gtable_0.2.0       rvest_0.3.2        psych_1.7.3.21
## [16] evaluate_0.10      inline_0.3.14      GGally_1.3.0
## [19] forcats_0.2.0      parallel_3.3.3     Rttf2pt1_1.3.4
## [22] broom_0.4.2        Rcpp_0.12.10       scales_0.4.1
## [25] jsonlite_1.3       mnormt_1.5-5       digest_0.6.12
## [28] hms_0.3            stringi_1.1.3      grid_3.3.3
## [31] tools_3.3.3        magrittr_1.5       lazyeval_0.2.0
## [34] extrafontdb_1.0    xml2_1.1.1         lubridate_1.6.0
```

```
## [37] assertthat_0.1     reshape_0.8.6      httr_1.2.1
## [40] R6_2.2.0           nlme_3.1-131
```