

Influence of Stadium and State Characteristics on Home Winning Percentage

Clara Livingston and Emily Kaegi

May 25, 2018

Introduction

Typically, when looking at statistics related to sports, researchers tend to focus on the dynamics of the team in question when predicting winning percentages. The goal of our research was to take analysis away from factors related to the team and investigate how factors relating to the location of the team's home stadium affects their home winning percentage. Most people are aware of the phrase "homefield advantage", but just how influential is a team's homefield in boosting their win percentages? SBNation and Sporting News both claim that home field advantage is real across most sports, especially the NBA. What appears harder to determine, however, is the cause of variability in this percentage between different sports and teams.

Instead of focusing on existence of "homefield advantage" we decided to look at what factors relating to a home stadium or arena might be correlated with winning percentage. Specifically, how do characteristics of the arena, franchise, fan base as well as state metrics affect a home team's winning. Are these characteristics consistent across all major league sports? We chose to focus on the NBA, NFL, MLB, and NHL since these professional leagues are the most popular in the United States.

Some of the most interesting potential predictors we wanted to explore were what percentage of a states population voted for Donald Trump in the 2016 presidential election. Does political leaning have any correlation with how well teams play at home? Also, using a metric of states' happiness level, we were curious if happiness is in anyway correlated with winning percentage. Do teams in happier states win more? While we cannot determine if this is casual or even the direction of the relationship (are states happy because their teams win or do players in happy states play better?) it is still interesting to explore.

Methods

Results

EDA Paragraphs

To discover the relationship between home game winning percentages and the variables of interest, we used logistic binomial regression. We observed variation in home win percentage by state, but did not want to include a variable for each state, we decided to include state as a random effect in our model.

Insert a couple sentences about modeling process description?

The heirarchical form of the final model is below. Estimates for the β parameters can be found in Table ??. Note that Sport Baseball is the baseline level.

Level 1 (team level):

$$\log odds(homewins) = a_i + \beta_0 Capacity_{ij} + \beta_1 Attendance_{ij} + \beta_2 SportBasektball_{ij} + \beta_3 SportFootball_{ij} + \beta_4 SportHockey_{ij} + \beta_5 FranchiseValue_{ij} + \beta_6 YearOpened_{ij} + \beta_7 Capacity_{ij} SportBasektball_{ij} + \beta_8 Capacity_{ij} SportFootball_{ij} + \beta_9 Capacity_{ij} SportHockey_{ij} + \beta_{10} Attendance_{ij} SportBasektball_{ij} + \beta_{11} Attendance_{ij} SportFootball_{ij} + \beta_{12} Attendance_{ij} SportHockey_{ij} + \beta_{13} Attendance_{ij} Capacity_{ij}$$

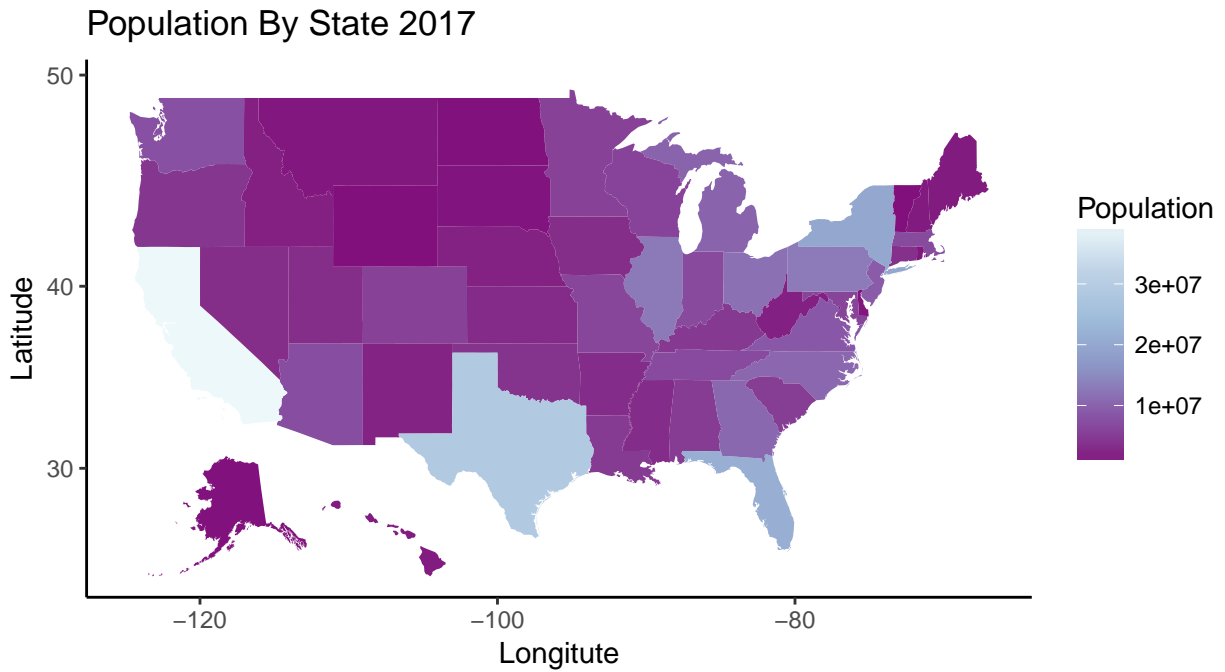


Figure 1: Caption Goes Here

Level 2 (state level):

$$a_i = \alpha_0 + \beta_{14}Population_i + \beta_{15}Happiness_i + \beta_{16}Happiness_iPopulation_i + u_i$$

Where $u_i \sim Norm(0, \sigma)$. Our model predicted $\hat{\sigma} = 0.1434$

Discussion

Appendix

Warning: Column `State` joining factors with different levels, coercing to character vector

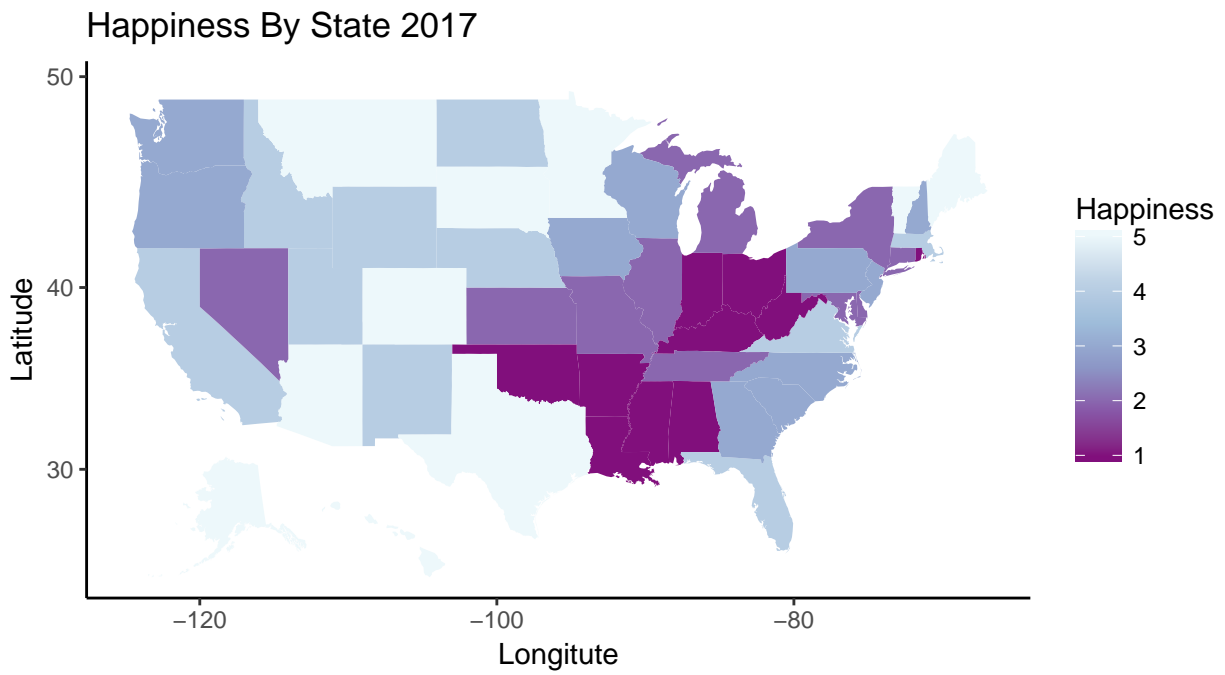


Figure 2: Caption Goes Here

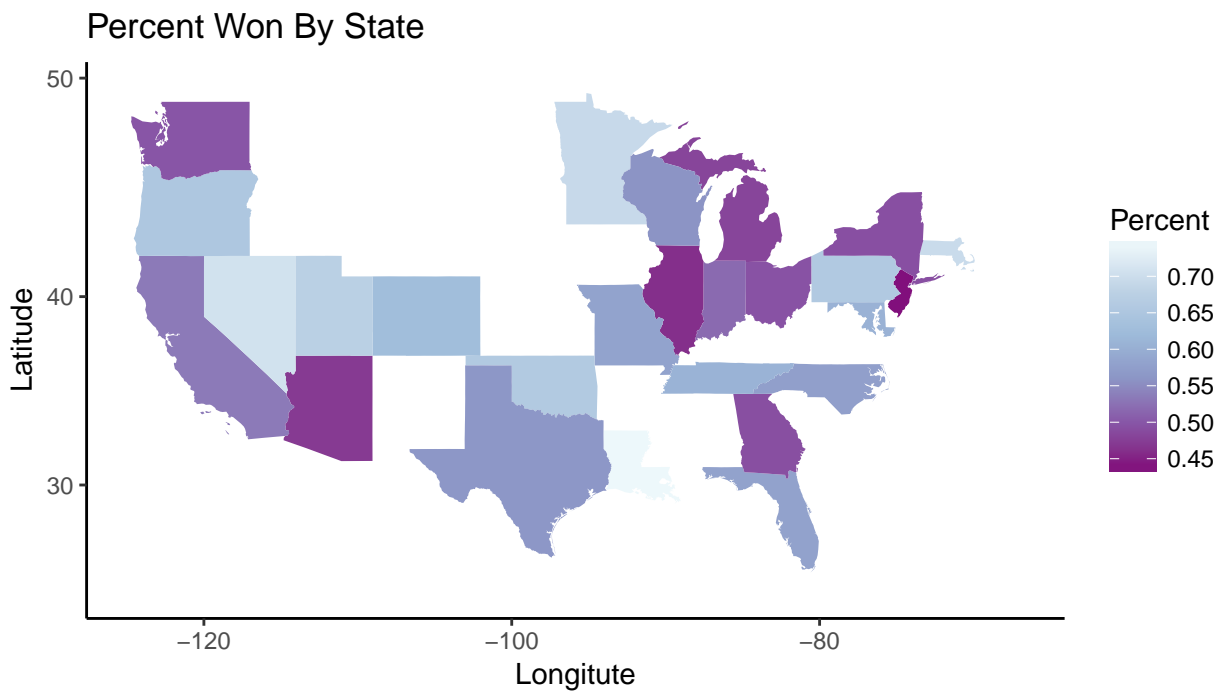


Figure 3: Caption Goes Here