The Perfect Fit

STAT 385 FA2018 - Team 24

Jacob Hart
Landon Hess
Jami Tugas
Joseph Wang

March 15, 2019

Abstract

This project examines possible variables that might affect median house value, along with several other variables such as climate, population density, job market, etc. This allows a user to find the most affordable area to live in while simultaneously satisfying their personal preferences of what type of environment they value.

Introduction

For our Project, we decided to make a program that makes it simple and easy for the user to find a county in the United States that fits their desired needs. When people are choosing where to live, there are some standard factors that most people focus on in order to choose the best county to live in (*Best Counties to Live in America* 2019).

With our program you will be able to customize the parameters of multiple factors, and immediately see a map of the results that match your selections. This program will use `R` and `shiny` in order to make an easy to use, and highly beneficial tool for people that are looking for counties with certain attributes. The data being used will be from multiple sources that has different information about each county in the United States. These factors will include things such as Average Temperature, Crime Rate, Population Density, etc. The user will be able to select parameters for each one in order to find the best county for them.

Related Work

There is a map of crime data (*CrimeMapping.com* 2019), a map of home prices (*Real Estate, Apartments, Mortgages & Home Values* 2019), and a website for climate data (*U.S. Climate Data* 2019), but there isn't one that combines all of these, and more.

Methods

When the program starts, web scraping will begin with `rvest` (Wickham 2016). The HTML of the page will have to be parsed into data frames. While this is happening, a loading screen will be displayed as this may take a while. Variables of interest will be spliced from the data sets and combined into a master data frame.

On the UI side, the main section will be a map of the counties of the United States, created with `ggplot2`. There will be a sidebar with two tabs. The first tab will be a search panel. The user will be able to select bounds for variables of interest (Avg. temperature, Crime rate, etc.), and search for counties that match these criteria. The counties on the map will be color coded based on how well they match the parameters. Counties that completely match the desired values will appear green. Those that have properties within some relative error will appear yellow, and the remaining counties will appear gray. From here, the user is able to select a county. This will automatically switch to the second tab, Details.

In the details tab, there will be bar graphs that show every variable of interest for the selected county, colored based on how well they fit the search criteria. These will also be created with `ggplot2`. Because of the interactivity of the project, `shiny` will of course be used.

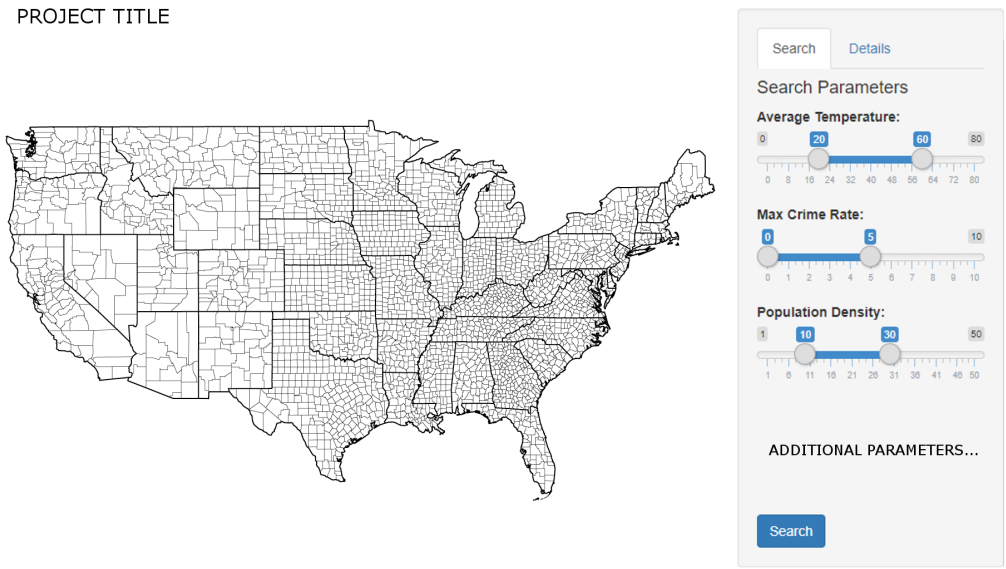

Figure 1: UI Prototype

Feasibility

- Is this project able to be completed before the end of the semester?
  - The concept that we chose for this project is doable and the applications needed in order to be successful are within our skill set. We thoroughly talked and envisioned what we wanted to include in our project. Furthermore, each member was assigned by a task that they're most comfortable with. With a clear vision, fair distribution of things to do, and established timeline, we're comfortable to say that we'll be able to complete this project by the end of the semester.
- What steps must occur to complete the project before the end of the semester?
  - Webscraping: Extract data from multiple websites, specifically data related to the crime rate, school system, temperature, median home value, etc. Essentially, factors that one might consider when considering the best county to live.
  - Parsing the data gathered, choosing only the variables we wanted to include, and combine these data to a big data frame.
  - Create the user interactive portion of the project
    - Create an interactive map of the United States, with a focus on every county.
    - Include a search panel and movable bounds for variables of interest
    - Include a details tab that shows exactly what the name suggests
  - Integrating the big data frame created from the parsed data into the UI portion
- What is the work plan to accomplish the necessary tasks before the end of the semester?
  - Specify who is doing what and when.
    - Webscarping: All members
    - Parsing and combining parsed data to a massive data frame: Jami
    - Interactive Portion: Landon, Joseph, Jacob
    - Data + UI: All members
  - Consider making a Gantt chart to highlight each stage of the project.

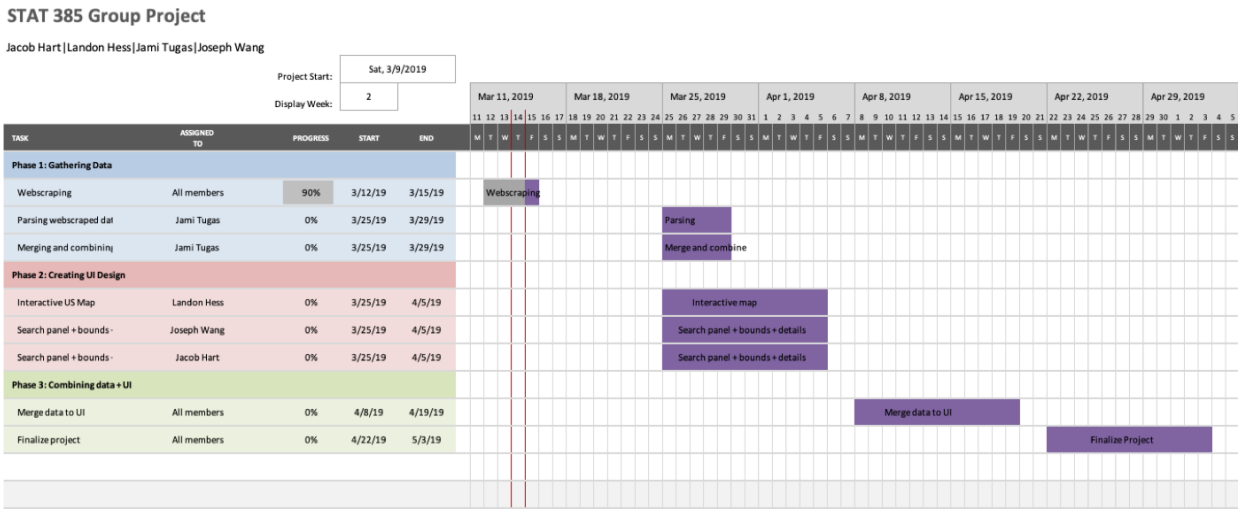

Figure 2: Expected Timeline

Conclusion

Essentially, this project is geared towards those looking for the most affordable location to live at given their own preferences. This is done by allowing users to filter out variables they don't care about. For example, someone who is already retired and doesn't have kids may not care a lot about the nearby education systems as opposed to a family with kids. Or an individual may prefer relatively colder weather as opposed to something like the deserts in Texas.

In summary, someone could pick what variable(s) they value most (such as a warmer climate or lower crime rate) and find the most affordable location based off of their own preference.

References

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2018). shiny: Web Application Framework for R. R package version 1.2.0. http://shiny.rstudio.com

Best Counties to Live in America. 2019. Niche.com Inc. https://www.niche.com/places-to-live/search/best-counties/.

CrimeMapping.com. 2019. Tritech Software Systems. https://crimemapping.com.

Real Estate, Apartments, Mortgages & Home Values. 2019. Zillow. https://zillow.com.

U.S. Climate Data. 2019. US Climate Data. https://www.usclimatedata.com/.

Wickham, Hadley. 2016. *Rvest: Easily Harvest (Scrape) Web Pages*. https://CRAN.R-project.org/package=rvest.