Lab 04 - La Quinta is Spanish for next to Denny's, Pt. 1

Visualizing spatial data

Your Name

Contents

Have you ever taken a road trip in the US and thought to yourself "I wonder what La Quinta means". Well, the late comedian Mitch Hedberg thinks it's Spanish for *next to Denny's*.

If you're not familiar with these two establishments, Denny's is a casual diner chain that is open 24 hours and La Quinta Inn and Suites is a hotel chain.

These two establishments tend to be clustered together, or at least this observation is a joke made famous by Mitch Hedberg. In this lab we explore the validity of this joke and along the way learn some more data wrangling and tips for visualizing spatial data.

The inspiration for this lab comes from a blog post by John Reiser on his *new jersey geographer* blog. You can read that analysis here. Reiser's blog post focuses on scraping data from Denny's and La Quinta Inn and Suites websites using Python. In this lab we focus on visualization and analysis of these data. However note that the data scraping was also done in R, and we we will discuss web scraping using R later in the course. But for now we focus on the data that has already been scraped and tidied for you.

1 Learning goals

- Visualising spatial data
- Joining data frames

2 Getting started

Go to the course GitHub organization and locate your homework repo, clone it in RStudio and open the R Markdown document. Knit the document to make sure it compiles without errors.

2.1 Warm up

Before we introduce the data, let's warm up with some simple exercises.

- Update the YAML, changing the author name to your name, and **knit** the document.
- Commit your changes with a meaningful commit message.
- Push your changes to GitHub.
- Go to your repo on GitHub and confirm that your changes are visible in your Rmd **and** md files. If anything is missing, commit and push again.

2.2 Packages

We'll use the **tidyverse** package for much of the data wrangling and visualisation and the data lives in the **dsbox** package. These packages are already installed for you. You can load them by running the following in your Console:

```
library(tidyverse)
library(dsbox)
```

2.3 Data

The datasets we'll use are called `dennys` and `laquinta` from the **dsbox** package. Note that these data were scraped from here and here, respectively.

Since the datasets are distributed with the package, we don't need to load them separately; they become available to us when we load the package. You can find out more about the datasets by inspecting their documentation, which you can access by running `?dennys` and `?laquinta` in the Console or using the Help menu in RStudio to search for `dennys` or `laquinta`. You can also find this information here and here.

To help with our analysis we will also use a dataset on US states, which is located in your repository's `data` folder.

```
states <- read_csv("data/states.csv")
```

Each observation in this dataset represents a state, including DC. Along with the name of the state we have the two-letter abbreviation and we have the geographic area of the state (in square miles).

3 Exercises

1. What are the dimensions of the Denny's dataset? (Hint: Use inline R code and functions like `nrow` and `ncol` to compose your answer.) What does each row in the dataset represent? What are the variables?

2. What are the dimensions of the La Quinta's dataset? What does each row in the dataset represent? What are the variables?

Knit, *commit, and push your changes to GitHub with an appropriate commit message. Make sure to commit and push all changed files so that your Git pane is cleared up afterwards.*

We would like to limit our analysis to Denny's and La Quinta locations in the United States.

3. Take a look at the websites that the data come from (linked above). Are there any La Quinta's locations outside of the US? If so, which countries? What about Denny's?

4. Now take a look at the data. What would be some ways of determining whether or not either establishment has any locations outside the US using just the data (and not the websites). Don't worry about whether you know how to implement this, just brainstorm some ideas. Write down at least one as your answer, but you're welcomed to write down a few options too.

We will determine whether or not the establishment has a location outside the US using the `state` variable in the `dennys` and `laquinta` datasets. We know exactly which states are in the US, and we have this information in the `states` dataframe we loaded.