Homework 5

DSO 545: Statistical Computing and Data Visualization

Spring 2018

Instructions

- Be sure to turn in a knit PDF or Word Document to Blackboard, **assignments turned in without this will not be graded**
- Make sure that when answering questions that your answer is displayed rather than saved in a variable (i.e. do not do things such as `answer <- {R Code}`)
- Make sure that all of your answers have code which produce them, answers given without the corresponding R code will not receive credit

Case 01

(1) Use `rvest` to scrape the table of the 100 most viewed YouTube videos from the following Wikipedia page (link: https://en.wikipedia.org/wiki/List_of_most-viewed_YouTube_videos).

(2) How many of the videos in the top 100 include Justin Bieber as the uploader / artist?

(3) Create an **EXACT** copy of the following graph. Use colors `magenta` and `dodgerblue`, and for the theme use `theme_bw()`.

Case 02: Los Angeles Collision

We have explored a bit of Los Angeles collisions data that goes from April 2008 to March 2013. This time let's plot these collisions on maps!

Questions

1. Setup a black and white map for LA. Save it to a variable called `LosAngeles`. Set your`maptype` to `roadmap` and the zooming parameter to `10`.

2. Reproduce the following map to show where the traffic collisions in LA were from 2008 till 2013? Choose an appropriate alpha (`0.01`), use color `dark blue`. Do you drive in LA? Describe what you see from the map you produced.

Santa Clarita
Cedar Springs
Hidden Springs
Angeles National Forest
Falling Sprin
San Fernando
La Cañada Flintridge
Burbank
Glendale
Pasadena
Azusa
Topanga State Park
Beverly Hills
Covina
West Covina
Los Angeles
East Los Angeles
City of Industry
Santa Monica
Inglewood
Downey
El Segundo
Brea
Manhattan Beach
Compton
Fullerton
Lakewood
Torrance
Anaheim
Long Beach
Garden Grove
Rancho Palos Verdes
Santa Ana
Google
Map data ©2018 Google

3. In traffic collision analysis, create a metric `KSI` (killed and severely injured) to identify the severeness of an accident. Create a new column `KSI`, put in `Yes` for those accidents with `HighestDegreeofInjury` of `Fatal` or `Severe Injury`. Mark `No` for other cases. Recreate the following map. (Use `size = 0.01`, `bins=100`, `alpha=0.5`, `red` color for Yes, and `black` color for No)

4. Now we focus on central Los Angeles areas and collisions with "KSI=YES" cases only. Reproduce a copy of the following graphs. (Use `bins=5`, and zooming parameter =12)