

PS6_{Gao}

Gaoningjing

march 2020

1 Question 3

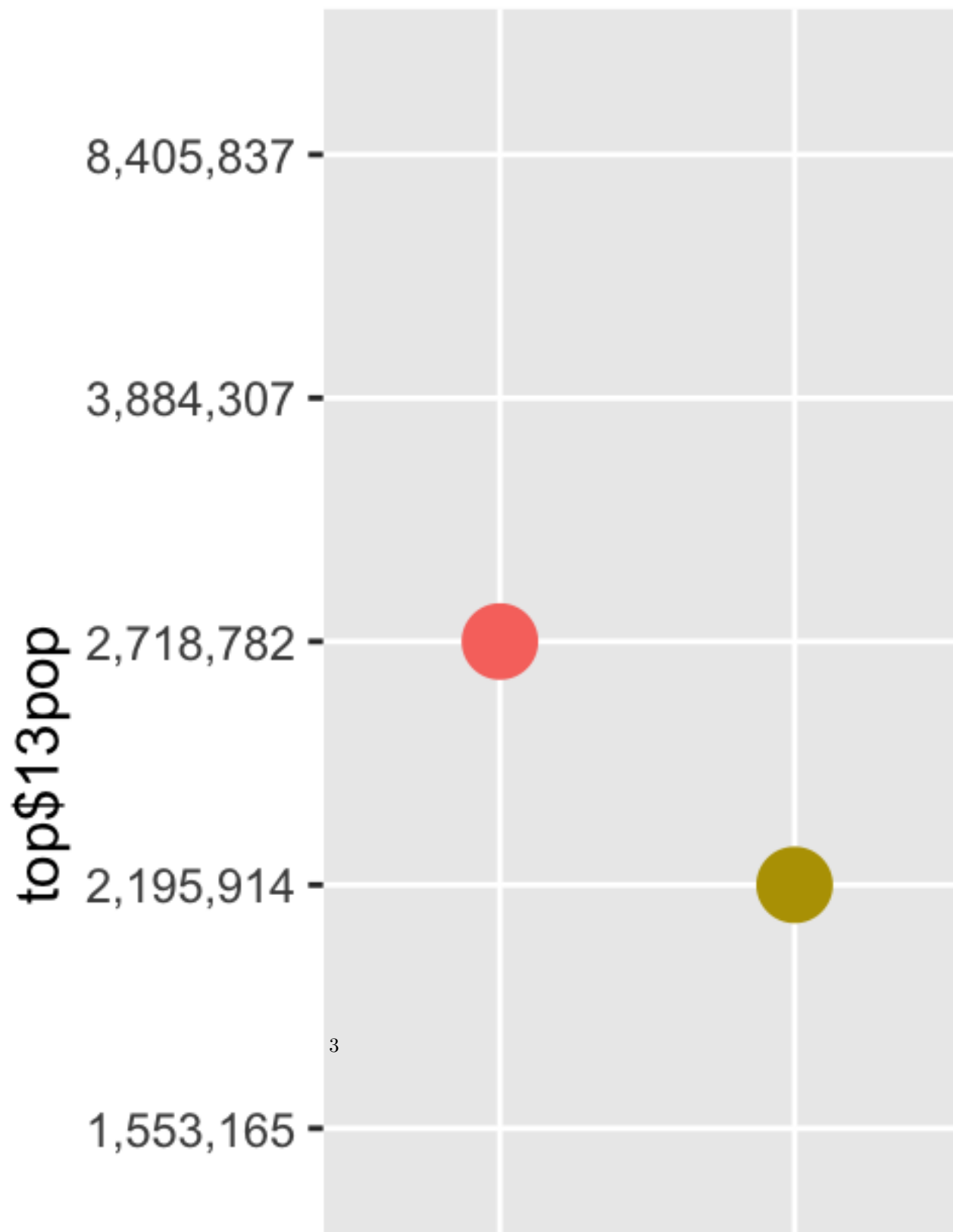
```
t library(twitteR) library(tm) library(wordcloud) library(RColorBrewer) library(SnowballC)
library(syuzhet) library(tidyverse) library(tidytext) library(ggplot2) library(forcats)
url <- paste("https://www.infoplease.com/us/us-cities/top-50-cities-us-population-
and-rank", sep="") html <- read_html(url) reading_the_html_code_into_memory_html_substr(html_text(html), 1, 1000)
html_table(html_str(tab)city <- tab[[1]]citynames(city)[which(names(city) ==
"SizerANK2014")] < -"rank"names(city)[which(names(city) == "7/1/2014POPULATIONESTIMATE")] <
-"14pop"names(city)[which(names(city) == "7/1/2013POPULATIONESTIMATE")] <
-"13pop"names(city)[which(names(city) == "4/1/2010CENSUSPOPULATION")] <
-"10pop"names(city)[which(names(city) == "7/1/2005POPULATIONESTIMATE")] <
-"05pop"names(city)[which(names(city) == "4/1/2000CENSUSPOPULATION")] <
-"00pop"names(city)[which(names(city) == "4/1/1990CENSUSPOPULATION")] <
-"90pop"city
```

2 Question 5

from my graphs, i see clearly know the rank of to 50 cities for 2014. it much clearer than dataset. also, i do the top 6 cities for 2013 and 2014 population. the gap of them are clear.



13 top6 cities



14 top6 cities

