largestCities

R Markdown

```
largestCity <- read.csv("data/gistfile1.txt",col.names = c("RANK","CITY","STATE","POP","GROWTH"))
largestCity$CITY <- tolower(largestCity$CITY)
largestCity$STATE  <- toupper(largestCity$STATE)
largestCity$RANK <- as.numeric(largestCity$RANK)
```

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
load("data/nih_2012-2016.dat")
nih_2016 <- nih %>%
  filter(year == 2016)
sel_nih_2016 <- nih_2016
sel_nih_2016$CITY <- tolower(sel_nih_2016$CITY)
sel_nih_2016 <- sel_nih_2016 %>%
  group_by(CITY,STATE = `STATE OR COUNTRY NAME`) %>%
  summarise(totalFunding = sum(FUNDING))
```

```
require(dplyr)
cityRanking_table <- inner_join(sel_nih_2016,largestCity, by = "CITY", convert = TRUE) %>%
  filter(STATE.x == STATE.y) %>%
  arrange(desc(STATE.x)) %>%
  group_by(STATE.x) %>%
  mutate(rank_funding = dense_rank(desc(totalFunding)), rank_size = dense_rank(RANK))


cityRanking_table <- cityRanking_table %>%
  select(STATE = STATE.x, CITY,pop_US = RANK,
         funding_state =  rank_funding, size_state =  rank_size, totalFunding) %>%
  arrange(desc(totalFunding))
cityRanking_table$funding_US <- order(desc(cityRanking_table$totalFunding))
cityRanking_table <- cityRanking_table %>%
  select(STATE, CITY, size_state, funding_state, pop_US, funding_US) %>%
  arrange(size_state)
write_csv(cityRanking_table, path = "data/cityRanking_table.csv")
```