# 1. Read the file into R

library(readr)

zillow\_median\_sales\_price\_per\_county <- read\_csv("Desktop/GitHub/Runjini\_Sprint\_6\_files/zillow\_median\_sales\_price\_per\_county.csv")

View(zillow\_median\_sales\_price\_per\_county)

# 2. Determine Column / Data information

length(zillow)

# 122 columns

max(zillow$SizeRank)

# 305

min(zillow$SizeRank)

# 1

# 3. Use R Studio to filter

# Filters work differently on numerical (ex. price) versus character (ex. state name) variables.

# 4. Using R Studio determine the range of the variables in each Particular Column

# Looking at filters on sale prices, the max has risen from the first dated column from $842K to over $1M

# 5. Utilize the "Unique()" command to determine variable names

unique(zillow$RegionName)

# Gives 271 unique values for the region

unique(zillow$StateName)

# Gives 38 unique states represented

# 6. Develop Ongoing questions

# What is the average sales price over time?

# What state has the highest average price overall? What state has the lowest average price overall?

# What is the correlation of size (assuming either physical or population) of a state versus its home sales prices?

# 7. Quick Analysis of distribution (barplot) of data

barplot(zillow$"2008-03", main="StateName")

barplot(zillow$"2008-03", main="RegionName")

1 CommentCollapse

Let me know if this doesn't work.