Homeless Data Analysis - Insights

Sudha Subramanian (Sparkfish)

Oct 28, 2017

Homelessness in United States - A Study

This is an analysis on a public dataset on homelessness across United States, downloaded from HUD website. For the scripts / datasets, please visit https://github.com/susub31/EARL-Boston2017-Presentation.

Analysis revealed that Hawaii had the highest homeless ratio (number of homeless people relative to the State's population) in the country. The number of people homeless in Hawaii may not seem like a high number compared to the number of homeless people in California or New York. Considering the percentage of people who are homeless relative to the population in the state, it paints an alarming picture as shown below.

Homeless Counts

"Point-in-time" surveys are taken periodically to track counts of homeless people periodically. This information is very useful to understand spread of homelessness across all States. It also feeds information to the Department of Housing and Urban Development to determine funding for cities and other initiatives that need to be taken to reduce homelessness.

```
# Read the dataset that holds Homeless Data for 2016
hdata <- read.csv("../Datasets/HomelessData2016.csv", stringsAsFactors = FALSE)

# Read the dataset that holds the extracted values of Lat and Lon for the CoCs
modhdata <- read.csv("../Datasets/HomelessData2016_With_LatLon.csv", stringsAsFactors = FALSE)
modhdata$Total.Homeless..2016 = NULL
modhdata$CoC.Name=NULL

# Merge the datasets and extract the desired columns; rename the columns
alldata <- merge(hdata, modhdata, by.x ="CoC.Number", by.y="CoC.Number")
alldata <- select(alldata, CoC.Number, CoC.Name, cityname, lat, lon, Total.Homeless..2016,
                  Homeless.Veterans..2016, Homeless.Unaccompanied.Youth..Under.25...2016) %>%
  rename(TotalHomeless = Total.Homeless..2016, HomelessVeterans = Homeless.Veterans..2016,
         HomelessYouth = Homeless.Unaccompanied.Youth..Under.25...2016)

# The following conversion is required to get the numeric value for the homeless counts
alldata$TotalHomeless <- as.numeric(gsub(",","", alldata$TotalHomeless))
alldata$HomelessVeterans <- as.numeric(gsub(",", "", alldata$HomelessVeterans))
alldata$HomelessYouth <- as.numeric(gsub(",", "", alldata$HomelessYouth))

# Extract the lat/lon coordinates pertaining to mainland US
# http://en.wikipedia.org/wiki/Extreme_points_of_the_United_States#Westernmost
top = 49.3457868 # north lat
bottom =  24.7433195 # south lat
left = -124.7844079 # west long
right = -66.9513812 # east long

#select for states in continous United States
contusdata <- alldata %>%
  filter(lat > 24 & lat < 50) %>%
```

```r
  filter(lon < -66 & lon > -124)

#Include a column to capture "State" from CoCNumber
contusdata$State = substr(contusdata$CoC.Number,1,2)

# Load US States dataset for generating the map
us <- map_data("state")
usstates <- read.csv("../Datasets/StateNames.csv")
usstates <- usstates %>%
  add_rownames("region") %>%
  mutate(region=tolower(StateName))

# Merge US States and Cost of Living by cities datasets for color-coding the US Map
usstates$StateName=NULL
usstates <- merge(usstates, contusdata, by.x="State", by.y="State")

# Group by and get totals by state in each category
CountsByState <- usstates %>%
  group_by(State, region) %>%
  summarize(TotalHomeless = sum(TotalHomeless), HomelessVeterans = sum(HomelessVeterans), HomelessYouth

CountsByState$Count.Category <- ifelse(CountsByState$TotalHomeless > 20000, "HIGH", ifelse(CountsBySta
```
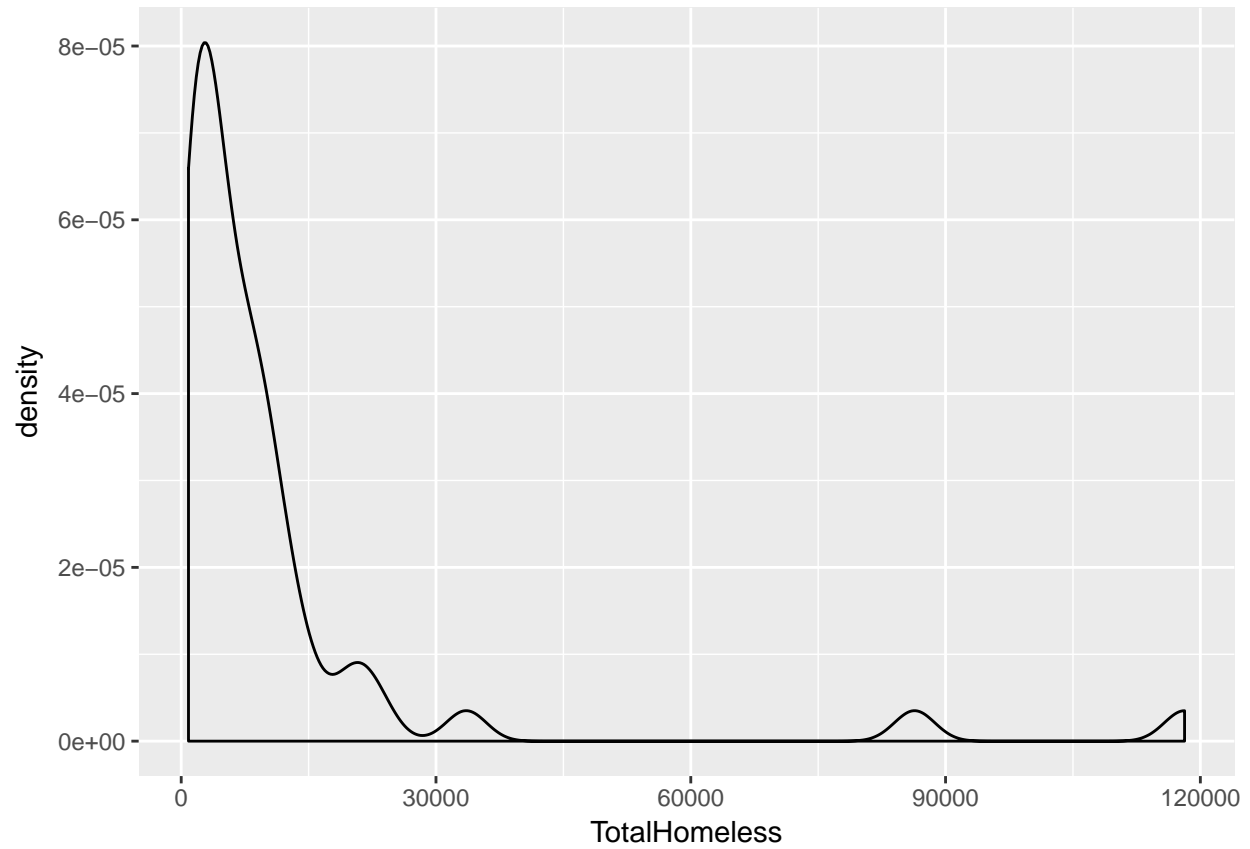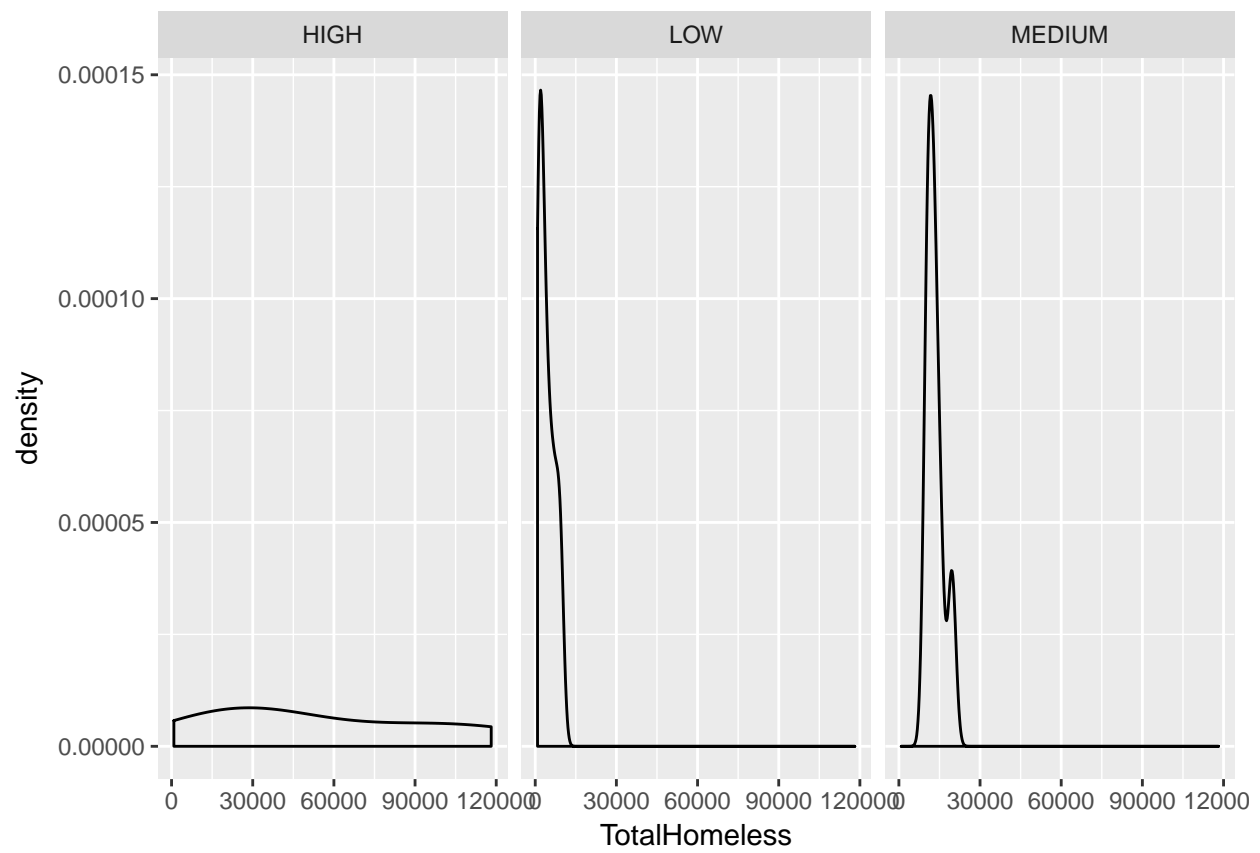
Density Map of Homeless Counts

```r
ggplot(CountsByState, aes(x=TotalHomeless)) +
  geom_density()
```

Density Map of Homeless Counts - by Category

```
ggplot(CountsByState, aes(x=TotalHomeless)) +
  geom_density() +
  facet_wrap(~ Count.Category)
```

Color-coded map of Homeless Counts

The homeless counts in the dataset is used to color code the US map to get an understanding of the extent of homelessness across different states. This will be used as the base map on which the counts in individual categories (Adults, Veterans, Youth) are plotted.

```r
#Plot a blank US Map
BlankUSMap <- ggplot()
BlankUSMap <- BlankUSMap + geom_map(data=us, map=us, aes(x=long, y=lat, map_id=region),
                                    fill="white", color="black")

#US Map, color coded based on the total homeless counts in the state
HomelessMap <- BlankUSMap + geom_map(data=CountsByState, map=us,
                              aes(fill=TotalHomeless, map_id=region), color="#ffffff", size=0.15)

HomelessMap <- HomelessMap + scale_fill_continuous(low='lightgrey', high='lightblue', guide='colorbar')
HomelessMap <- HomelessMap + labs(x=NULL, y=NULL) +
  theme(panel.border = element_blank()) +
  theme(panel.background = element_blank()) +
  theme(axis.ticks = element_blank()) +
  theme(axis.text = element_blank())

HomelessMap
```

Homeless Counts (Adults) - Top 100
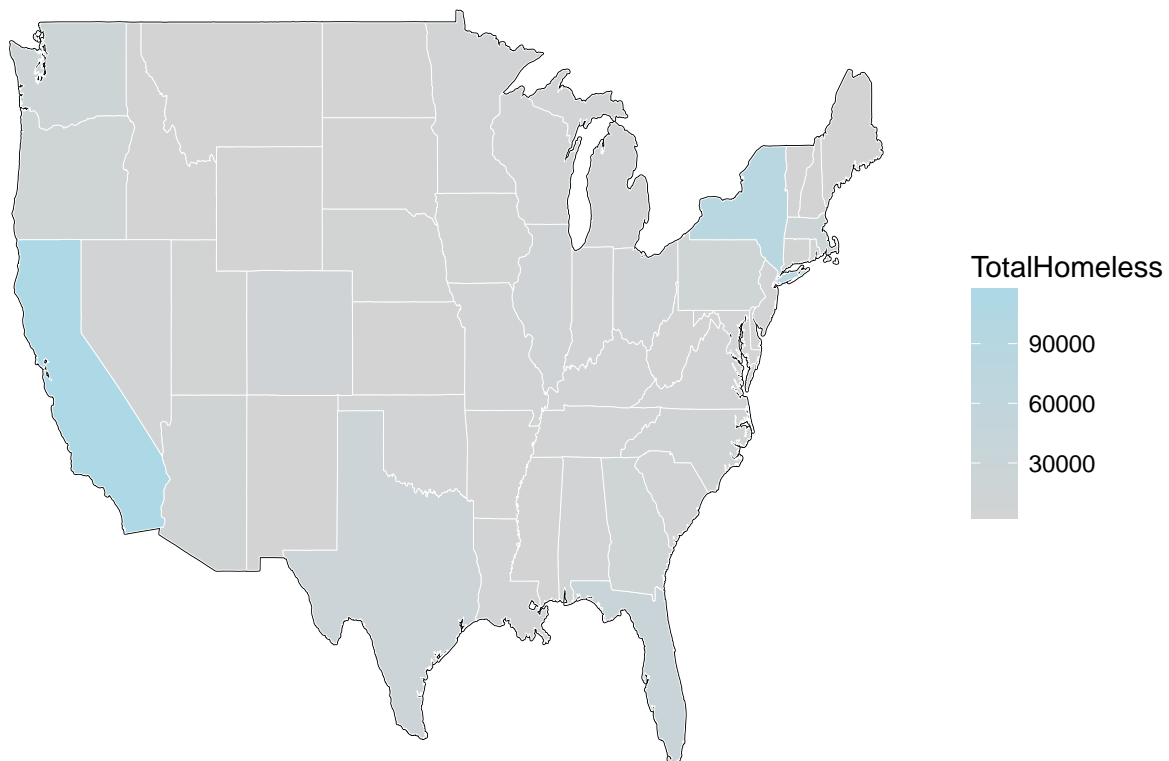
Plotting the count of homeless people (adults) for the continuums (CoC) with high counts of
homelessness, paints the following picture, where size of the bubble indicates high count of
homelessness.

```
# select top 100 based on homeless counts - adults
hdatatop100 <- tbl_df(subset(alldata, !((substr(alldata$CoC.Number,1,2) =="HI")) & !((substr(alldata$CoC
  top_n(100, TotalHomeless)

Top100HomelessAdultsMap <- HomelessMap +
  geom_point(aes(x=lon, y=lat, size=TotalHomeless, colour="red", alpha=0.8), data=hdatatop100) +
  ggtitle("Homeless Counts - Adults (Top 100)") +
  scale_size_continuous(name="Homeless Counts", range = c(2,12), guide = FALSE) +
  scale_alpha(guide=FALSE) +
  scale_colour_discrete(guide=FALSE)

Top100HomelessAdultsMap
```


Homeless Counts – Adults (Top 100)

Homeless Counts (Veterans) - Top 100

Plotting the count of homeless people (veterans) for the continuums (CoC) with high counts
of homelessness, paints the following picture, where size of the bubble indicates high count of
homelessness.

```r
# select top 100 based on homeless counts - veterans
hvetsdatatop100 <- tbl_df(subset(alldata, !((substr(alldata$CoC.Number,1,2) =="HI")) & !((substr(alldata
  top_n(100, HomelessVeterans)

Top100HomelessVeteransMap <- HomelessMap +
  geom_point(aes(x=lon, y=lat, size=HomelessVeterans, colour="red", alpha=0.8), data=hvetsdatatop100) +
  ggtitle("Homeless Counts - Veterans (Top 100)") +
  scale_size_continuous(name="Homeless Counts", range = c(2,12), guide = FALSE) +
  scale_alpha(guide=FALSE) +
  scale_colour_discrete(guide=FALSE)

Top100HomelessVeteransMap
```

Homeless Counts – Veterans (Top 100)

**Homeless Counts (Youth) - Top 100**

Plotting the count of homeless people (Youth) for the continuums (CoC) with high counts of homelessness, paints the following picture, where size of the bubble indicates high count of homelessness.

```
hyouthdatatop100 <- tbl_df(subset(alldata, !((substr(alldata$CoC.Number,1,2) =="HI")) & !((substr(alldat
  top_n(100, HomelessYouth)

Top100HomelessYouthMap <- HomelessMap +
  geom_point(aes(x=lon, y=lat, size=HomelessYouth, colour="red", alpha=0.8), data=hyouthdatatop100) +
  ggtitle("Homeless Counts - Youth (Top 100)") +
  scale_size_continuous(name="Homeless Counts", range = c(2,12), guide = FALSE) +
  scale_alpha(guide=FALSE) +
  scale_colour_discrete(guide=FALSE)

Top100HomelessYouthMap
```


Homeless Counts – Youth (Top 100)
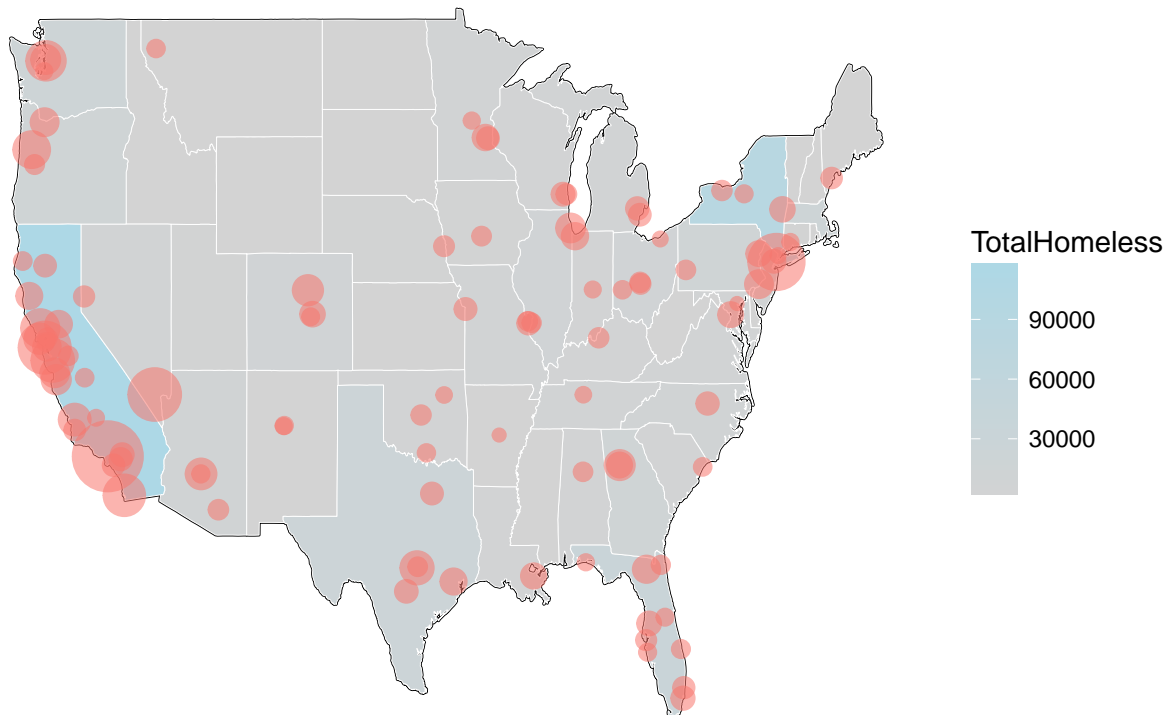
Grants vs. Homeless Counts - An analysis

Read ESG Grants Dataset (with geocodes for mapping); exclude mapping for Hawaii, Puerto Rico so that the focus is on mainland United States only. Extract the data pertaining to top 'n' (in this example, top **25** are selected) for both the homeless counts dataset and grants dataset for performing a comparative analysis. As can be seen from the map, there is disparity between where we see top homelessness and where top grants are allocated.

For the base map, the cost of living index values are used to color-code the US map. Thus, it is possible to see how the homelessness correlates to the cost of living index values

```r
#Read Grants dataset
GrantsDS <- read.csv("../Datasets/ESGGrantDSWithGeoCodes.csv")
GrantsDStop35 <- tbl_df(GrantsDS) %>%
  top_n(35, TOT_AMT)
GrantsDStop35 <- subset(GrantsDStop35, !(STUSAB == "PR"))

AllGrantsDStop35 <- group_by(GrantsDS, round(lat, 2)) %>% mutate(sum(TOT_AMT))

AllHomelessTop35 <- tbl_df(subset(alldata, !((substr(alldata$CoC.Number,1,2) =="HI")) & !((substr(alldat
  top_n(25, TotalHomeless)
GrantsDStop35 <- tbl_df(subset(GrantsDS, !((substr(GrantsDS$STUSAB,1,2) =="HI")) & !((substr(GrantsDS$ST
  top_n(25, TOT_AMT)


us <- map_data("state")
usstates <- read.csv("../Datasets/StateNames.csv")
usstates <- usstates %>%
  add_rownames("region") %>%
  mutate(region=tolower(StateName))

COLByCities <- read.csv("../Datasets/COL_ByCities.csv")
COLByCities$UrbanArea <- gsub(", ", " ", COLByCities$UrbanArea)
COLByCities$State <- COLByCities$UrbanArea
COLByCities$State <- substrRight(COLByCities$State, 2)

usstates$StateName=NULL
usstates <- merge(usstates, COLByCities, by.x="State", by.y="State")

COLByCities <- group_by(COLByCities, State)
COLByCities <- summarise(COLByCities, COLStateAvg=mean(COLIndex), HousingIndex=mean(Housing))

COLMap <- BlankUSMap + geom_map(data=usstates, map=us,
                                aes(fill=Housing, map_id=region), color="#ffffff", size=0.15)

COLMap <- COLMap + scale_fill_continuous(low='lightgray', high='lightblue', guide='colorbar')

COLMap <- COLMap + labs(x=NULL, y=NULL) +
  theme(panel.border = element_blank()) +
  theme(panel.background = element_blank()) +
  theme(axis.ticks = element_blank()) +
  theme(axis.text = element_blank())

GrantsAndCountsmap <- COLMap +
```

```
    geom_point(aes(x=lon, y=lat, size=TOT_AMT, colour="Total Grants", alpha=0.8), data=GrantsDStop35) +
    geom_point(aes(x=lon, y=lat, size=TotalHomeless, colour="Homeless Counts", alpha=0.8), data=AllHomeles
    labs(x='Longitude', y='Latitude') +
    ggtitle("Homeless Counts vs. Grants") +
    scale_alpha(guide=FALSE) +
    theme(legend.position="none")

GrantsAndCountsmap <- GrantsAndCountsmap + labs(x=NULL, y=NULL) +
    theme(panel.border = element_blank()) +
    theme(panel.background = element_blank()) +
    theme(axis.ticks = element_blank()) +
    theme(axis.text = element_blank()) +
    theme(legend.position = "none")

GrantsAndCountsmap
```

Homeless Counts vs. Grants