

# EARL-Houston-Pres

*Sudha Subramanian*

*November 14, 2018*

## Homelessness - Visualizations using R

Homelessness is a social problem at large and let's try to visualize the extent of homelessness in the United States. A recent article published in LA Times shows that there are about 120,000+ people homeless in California, of which about 55,000 are homeless in LA and its suburbs.

Let's visualize the homeless counts across all states, normalized to the population in the state.

```
# US Map of homeless counts
hdata <- read.csv("../Datasets/HomelessData2016_With_LatLon.csv")
hdata_2016 <- read.csv("../Datasets/PIT_CountsByState_2016.csv")
hdatatop25 <- tbl_df(hdata_2016) %>%
  top_n(25, TotalHomeless2016)

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

# Merge US States and Homeless counts (2016) datasets for color-coding the US Map
usstates$StateName=NULL
usstates <- merge(usstates, hdata_2016, by.x="State", by.y="State")

#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 COL Index in the state
```

## PIT (Point in Time) counts

PIT counts refer to the number identified as homeless on a specific day in January, which is used for various reporting on Housing and Urban Development (HUD) website. This data is used for the purpose of this analysis.

```
all_homeless <- read.csv("../Datasets/PITCounts_Consolidated_2013_to_2017.csv")

usstates <- read.csv("../Datasets/StateNames.csv")
usstates <- usstates %>%
  tibble::rownames_to_column("region") %>%
  mutate(region=tolower(StateName))

# Merge US States and Homeless counts datasets for color-coding the US Map
US_Homeless <- usstates
```

```

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

hpop <- read.csv("../Datasets/Homeless_by_Population.csv")

usstates <- read.csv("../Datasets/StateNames.csv")
usstates <- usstates %>%
  tibble::rownames_to_column("region") %>%
  mutate(region=tolower(StateName))

# Merge US States and Homeless counts datasets for color-coding the US Map
US_Homeless <- usstates
US_Homeless$StateName=NULL
US_Homeless <- merge(US_Homeless, hpop, by.x="State", by.y="State")

#US Map, color coded based on the COL Index in the state
HMap <- BlankUSMap + geom_map(data=US_Homeless, map=us,
                              aes(fill=PctHomeless2017, map_id=region), color="#ffffff", size=0.15)

HMap <- HMap +
  scale_fill_continuous(low='gray94', high='gray10', guide='colorbar') +
  theme(legend.position = "none")

HMap <- HMap + 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())

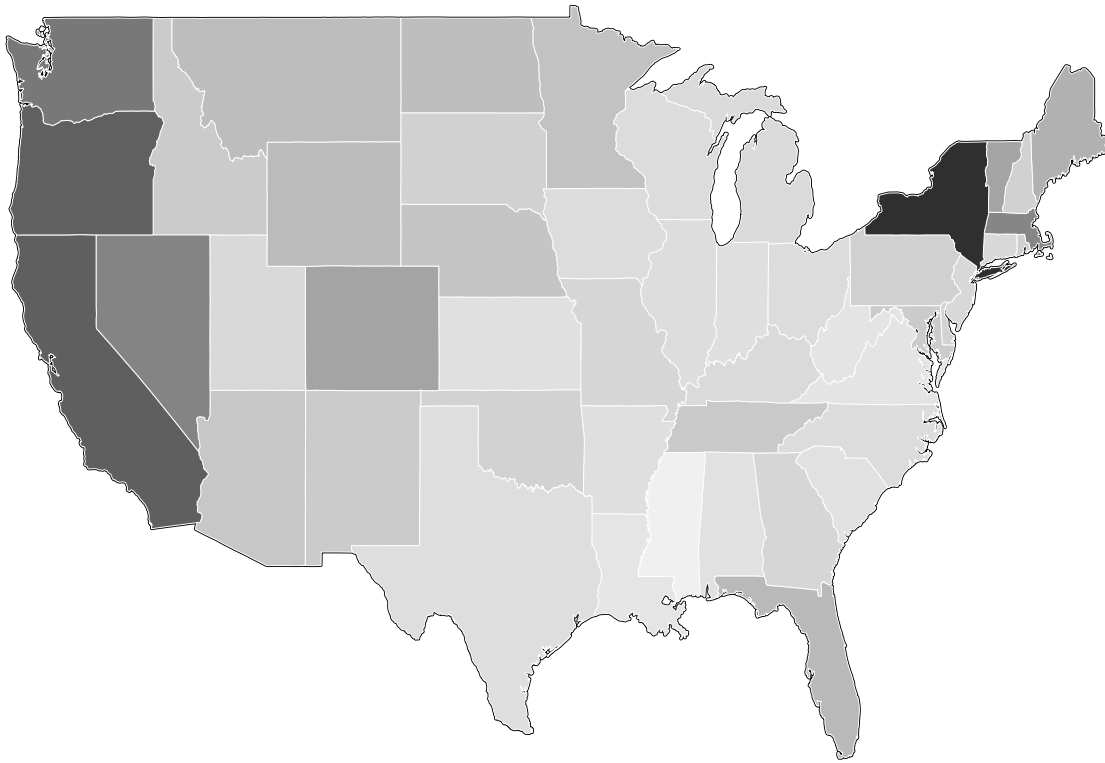
ggthemr('flat', type="outer", layout='scientific', spacing=2)

## Warning: New theme missing the following elements: panel.grid, plot.tag,
## plot.tag.position

ggthemr_reset()

print(HMap)

```



## Change in percent sheltered

HUD records information on the number of homeless people, number of sheltered vs. unsheltered and various other categories. Let's take a look at the number of sheltered homeless and its change over the years 2007 to 2017.

Let's visualize how sheltered counts have trended over the past few years in a few states. New York, although it shows a flat line, the actual bed counts in shelters went up drastically in 2016. Prior to that, people were housed in run-down motels or some emergency shelters.

```
pctsh <- read.csv("../Datasets/PctSH-2007-to-2017.csv")

theme_set(theme_bw())

sub1 <- transpose(pctsh[, -1])
colnames(sub1) <- pctsh$State
sub1$Year = c(2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007)

sub2 <- sub1 %>% select(Year, 'CA', 'NY', 'WA', 'FL')
sub3 <- tidyr::gather(sub2, "State", "Sheltered", 2:5)
sub4 <- sub3 %>% filter(State %in% c('CA', 'NY', 'WA', 'FL'))

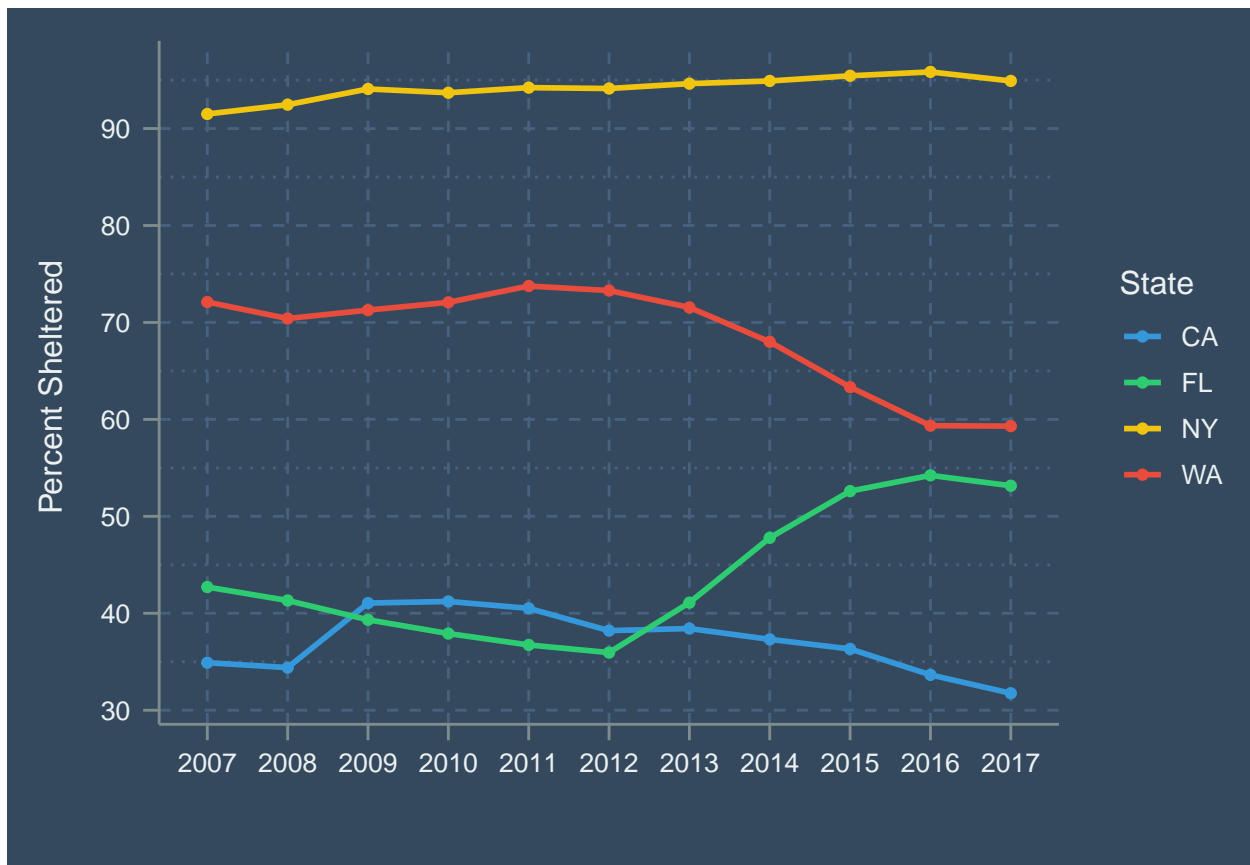
gg2 <- ggplot(sub4, aes(as.factor(Year), Sheltered, group=State, color=State)) +
  geom_point() + geom_line(size=1) +
  xlab("") + ylab("Percent Sheltered ") +
  scale_y_continuous(breaks=seq(30,100,by=10))
```

```
# theme_classic() #+ theme(legend.position = "none")

ggthemr('flat dark', type="outer", layout='scientific', spacing=2)

## Warning: New theme missing the following elements: panel.grid, plot.tag,
## plot.tag.position
#ggthemr('earth', type="outer", layout='scientific', spacing=2)

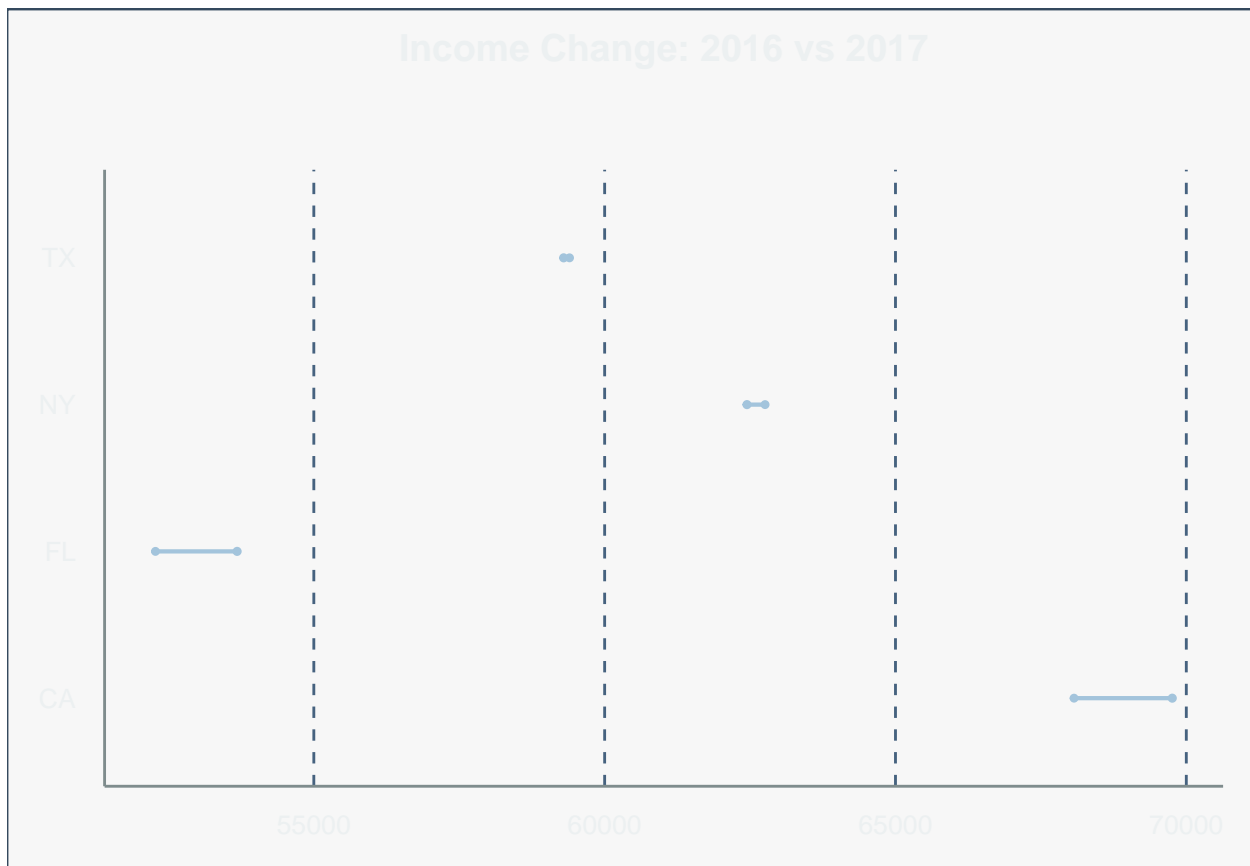
gg2
```



/newpage

## Change in INCOME

Income plays a major role in how an individual is able to survive. As we all know, cost of living keeps changing, more increasing across years and so, let's try to visualize how income has changed in these states in the recent year compared to the previous one.



```
usstates <- read.csv("../Datasets/StateNames.csv")
usstates <- usstates %>%
  tibble::rownames_to_column("region") %>%
  mutate(region=tolower(StateName))

# Merge US States and Homeless counts (2016) datasets for color-coding the US Map
US_Income <- usstates
US_Income$StateName=NULL
US_Income <- merge(US_Income, med_income, by.x="State", by.y="StateName")

cyear <- 2017
#US Map, color coded based on the COL Index in the state
IMap <- BlankUSMap + geom_map(data=US_Income, map=us,
  aes(fill=`2015`, map_id=region), color="#ffffff", size=0.15)

IMap <- IMap + scale_fill_continuous(low='gray94', high='light blue', guide='colorbar')

IMap <- IMap + 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())

HByYear <- med_income
newcolnames <- c("Y2010", "Y2011", "Y2012", "Y2013", "Y2014", "Y2015", "Y2016", "Y2017", "StateName")
colnames(HByYear) <- newcolnames
```

```

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

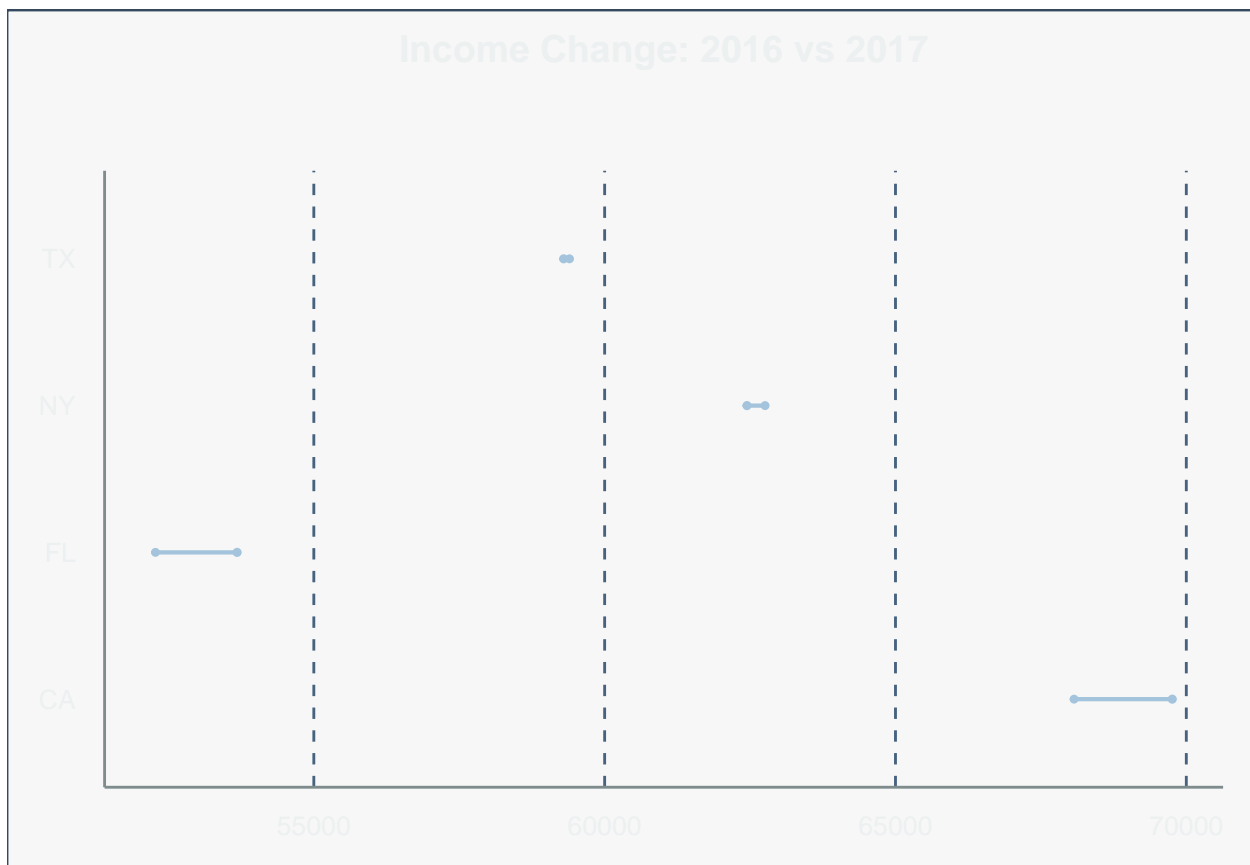
for (i in 1:(length(newcolnames)-1)) {
  IMap <- BlankUSMap + geom_map(data=US_Homeless, map=us,
                                aes(fill=newcolnames[i], map_id=region), color="#ffffff", size=0.1)

  IMap <- IMap + scale_fill_continuous(low='gray94', high='light blue', guide='colorbar')

  IMap <- IMap + 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())

  IMap
  IMap
}

```



## Homeless Counts - 2013 to 2017

Let's try to understand how the actual counts have varied across 2013 to 2017 in the states of interest. Florida, by adopting 'Housing First' policy has shown to reduce the number of homeless counts considerably. However, states such as California, New York and Washington see increase in the counts. So, let's visualize how the counts have changed.

```
h2017_top10 <- tbl_df(all_homeless) %>%
  top_n(7, HTotal_2017) %>%
  arrange(desc(HTotal_2017))

htotal_top10 <- tbl_df(all_homeless) %>%
  top_n(8, HTotal_2017) %>%
  select(State, HTotal_2017, HTotal_2016, HTotal_2015, HTotal_2014, HTotal_2013) %>%
  arrange(desc(HTotal_2017))

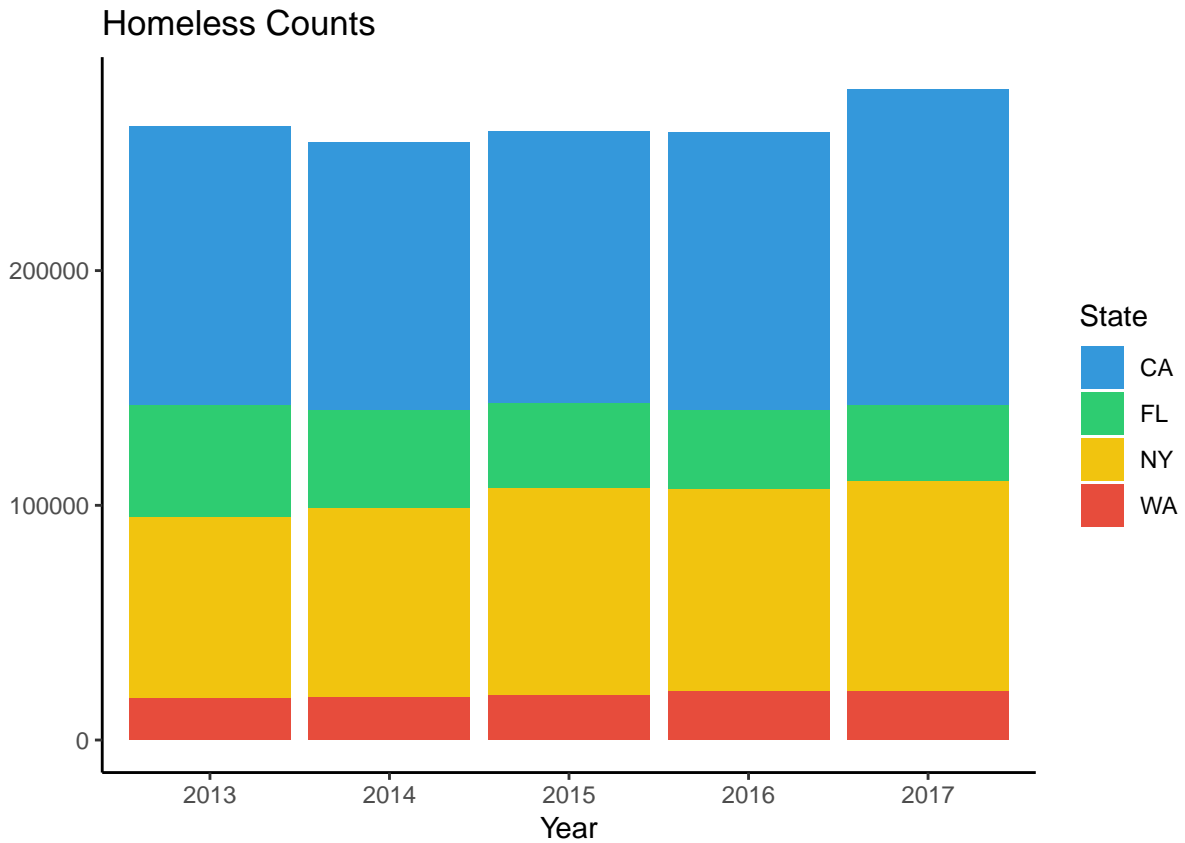
h <- tidyr::gather(htotal_top10, "HTotal_Year", "TotalHomeless", 2:6)
h$Year <- substrRight(h$HTotal_Year, 4)

p <- ggplot(data=h, aes(x=Year, y=TotalHomeless, color=State)) +
  geom_point(size=3) +
  ylim(10000, 150000)

#p <- ggplot(data=d1, aes(x=Year, y=TotalHomeless, color=State)) + geom_line() + geom_point()
#p <- ggplotly(p)

#format(y, scientific = FALSE)
p2 <- ggplot(data=h %>% filter(State %in% c("CA", "FL", "NY", "WA")), aes(x=Year)) +
  geom_bar(aes(y=TotalHomeless, fill=State), stat = "identity") +
  labs(title="Homeless Counts", y="", x = "Year") +
  #ylim(0, 300) +
  #scale_y_continuous(labels=function(n) {format(n, scientific = FALSE)}) )
  #expand_limits(y=0) + #axis(labels=format(scientific=FALSE)) +
  theme_classic() #+ theme(legend.position = "none")

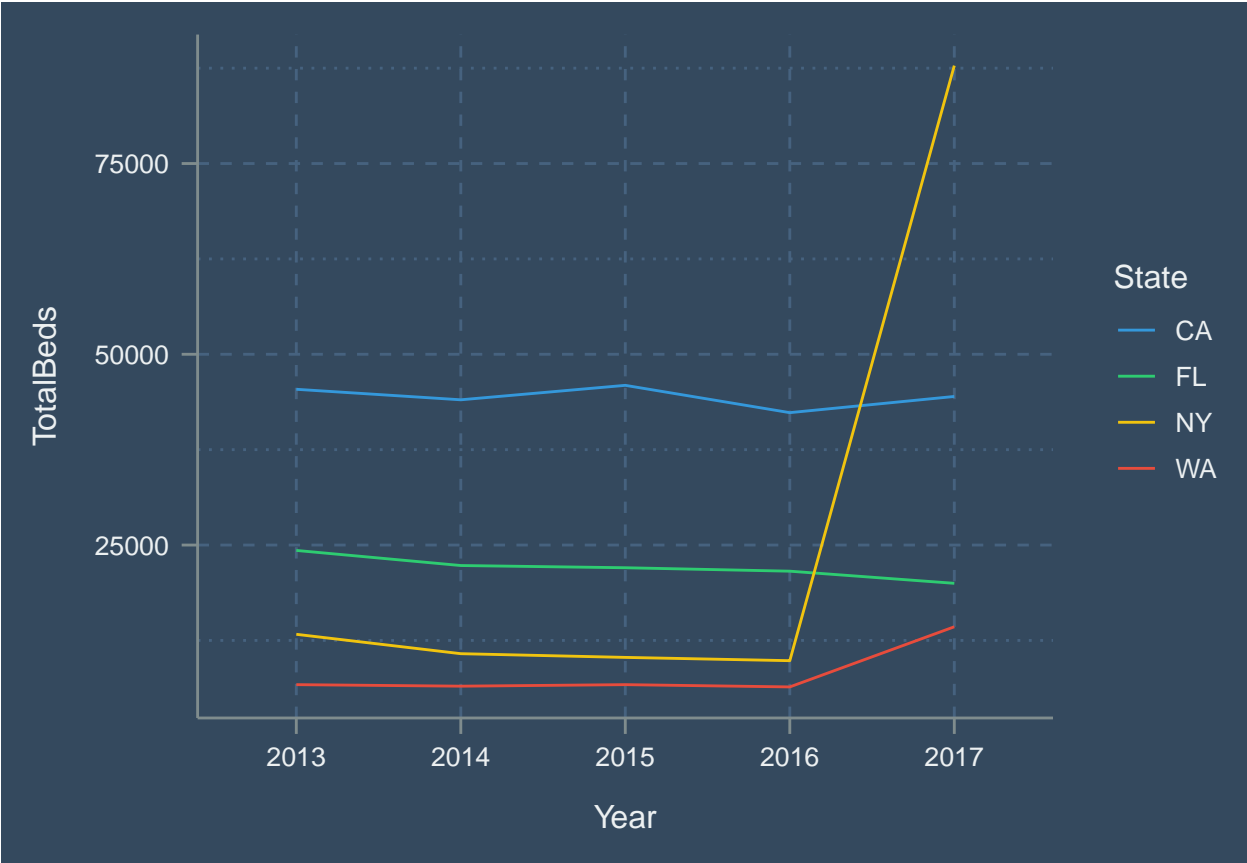
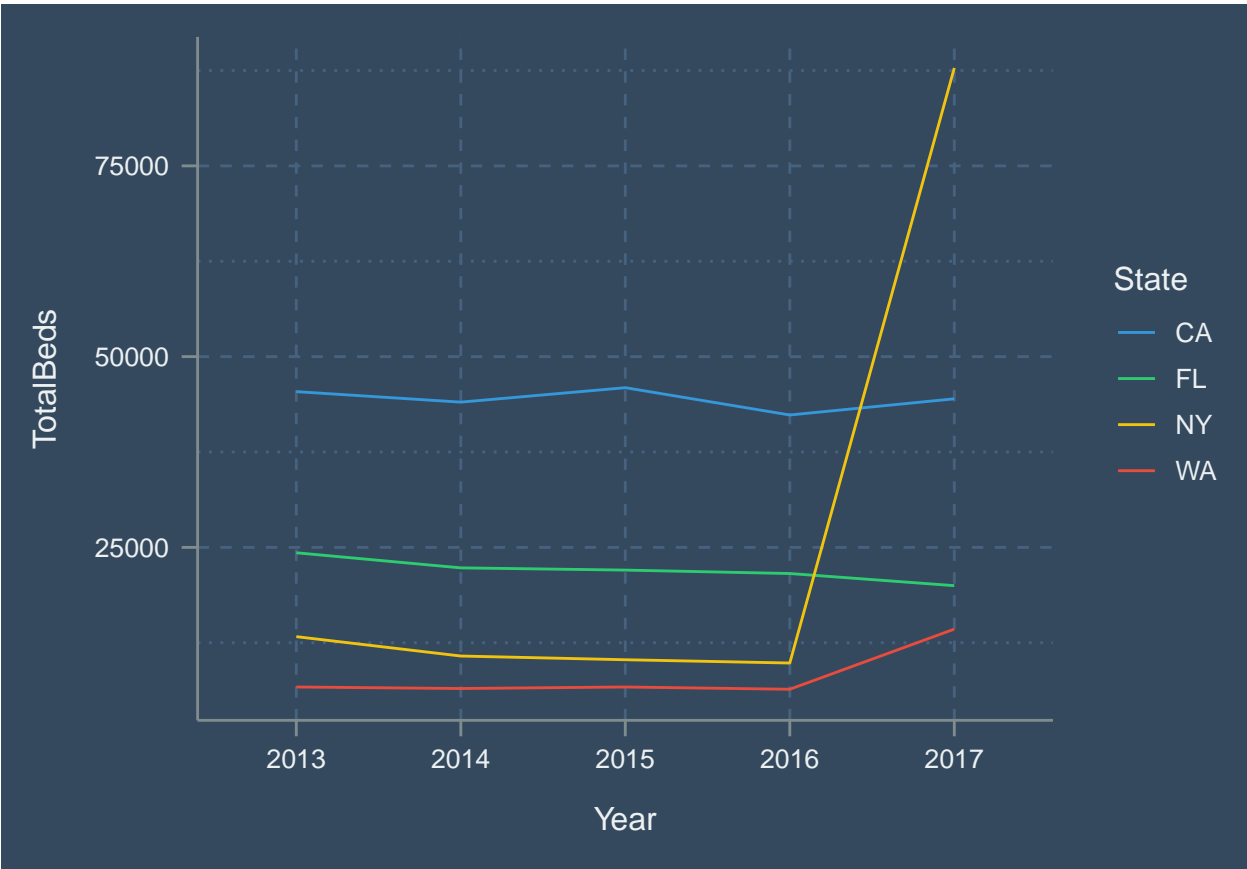
p2
```

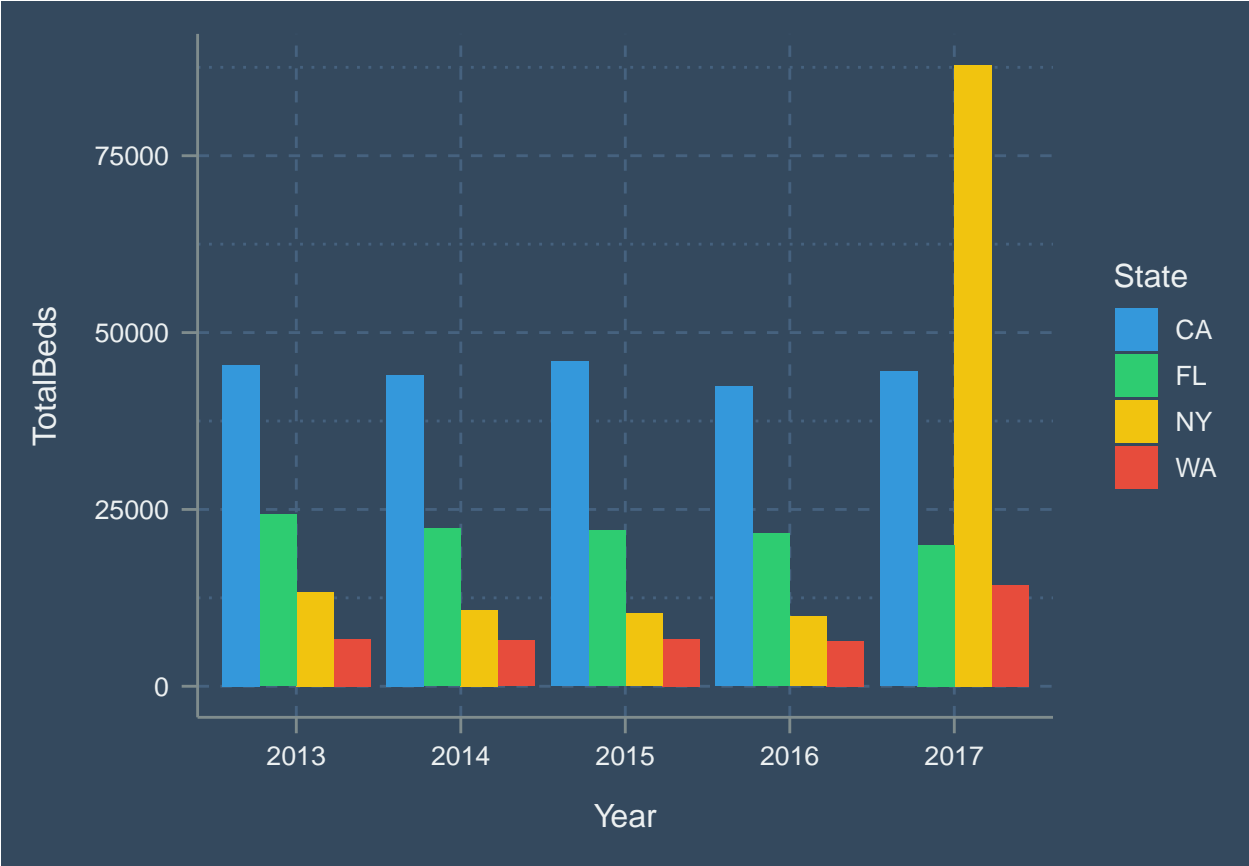


### Change in Bed Counts

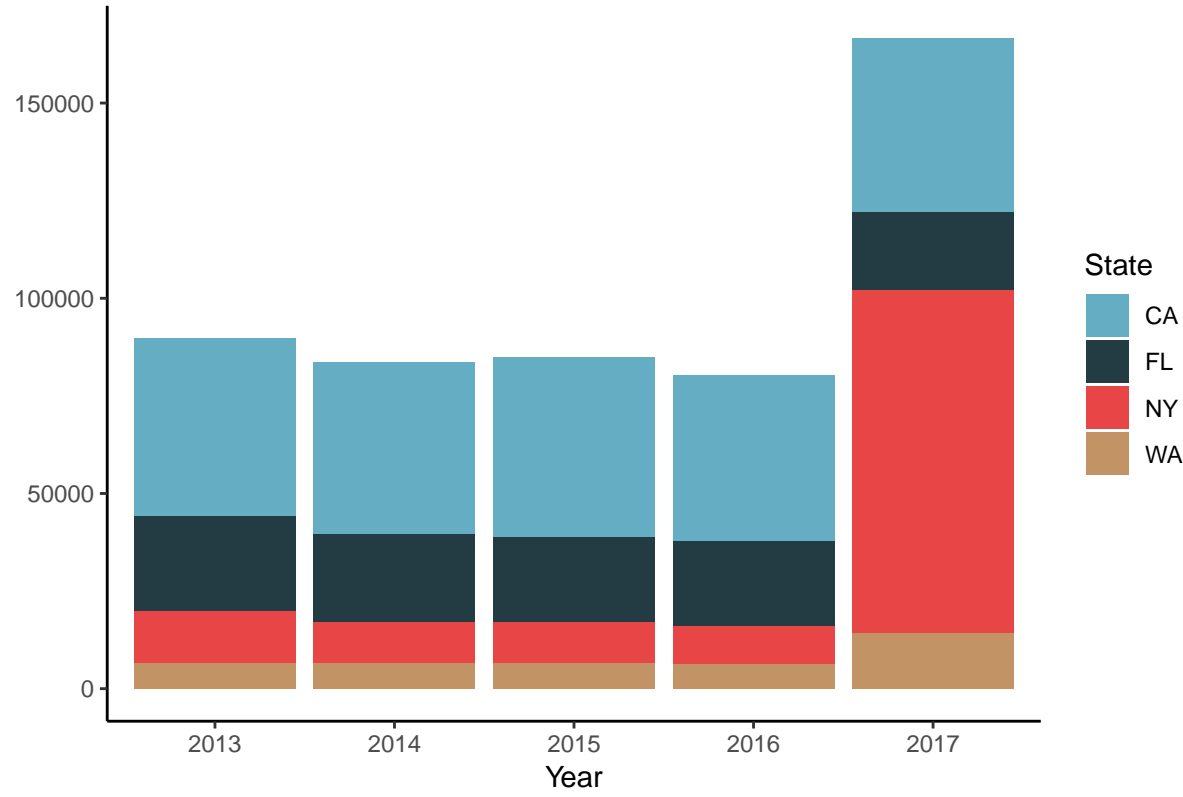
One of the key things many states have adopted is to provide shelters / beds for unsheltered homeless so that people don't have to spend nights out on the streets. They have also experimenting by providing affordable housing as a long-term solution. New York has built new shelters and has increased the bed counts multi-fold in 2016, as shown in the visualization below.







Shelter – Bed Counts



## Polar Plot

The visualization below is an interesting way of showing how the bed counts have changed in a radial fashion. As seen here, New York has seen considerable increase in bed counts between 2016 and 2017, while it is almost same for California and Washington. Florida shows a reduced number of bed counts for 2017 compared to previous years.

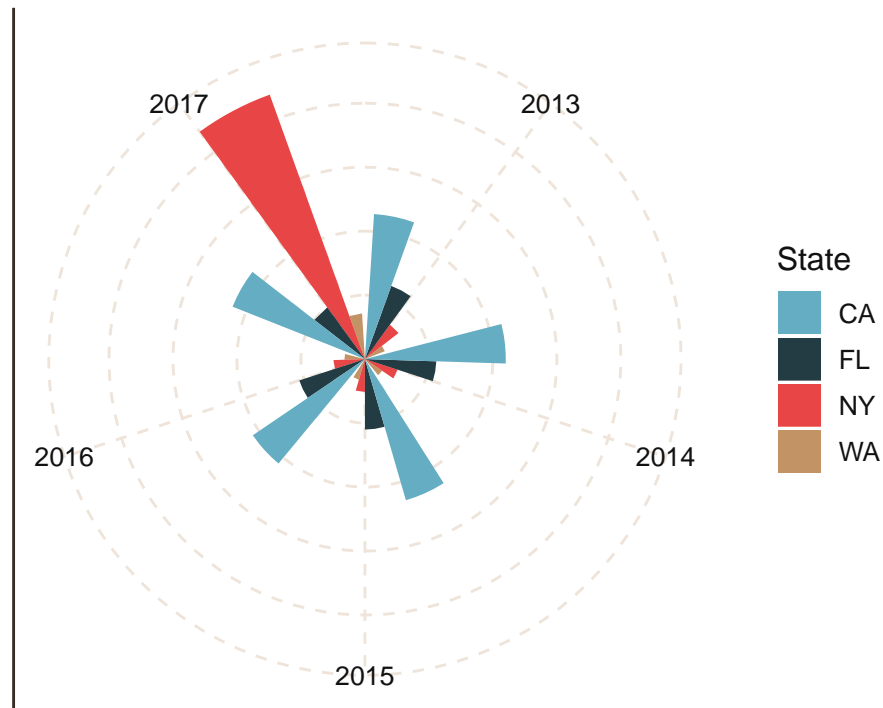
```
# make bar charts instead of stacked column charts
h <- ggplot(hic, aes(Year, TotalBeds))
h <- h + geom_bar(stat="identity", aes(fill=State), position="dodge")

h <- h + theme(axis.title.y=element_blank(),
               axis.text.y=element_blank(),
               axis.ticks.y=element_blank()) + xlab("") + ylab("")

h <- h + coord_polar()

h2 <- h + ylab("") + labs(title="Shelter - Bed Counts")
h2 <- h2 + ylim(0, 250000)

h
```



## Change in Chronically Homeless Counts

Chronically homeless refers to those who have been homeless for a prolonged period or have been in emergency shelters for a long time. Let's try to see how this has changed over the years as a divergent chart.

```
ch_data <- read.csv("../Datasets/ChronicallyHomelessDataset.csv")

#theme_set(theme_bw())
# Data Prep
#First subset the dataframe to the the dataframe of interest
ch_sub <- subset(ch_data, select = c("State", "PctChange_2007_2017"))

ch_sub$NormPct <- round(ch_sub$PctChange_2007_2017 - median(ch_sub$PctChange_2007_2017), 2)
ch_sub$NormDir <- ifelse(ch_sub$NormPct < 0, "below", "above")
ch_sub <- ch_sub[order(ch_sub$NormPct), ]
ch_sub$State <- factor(ch_sub$State, levels=ch_sub$State)

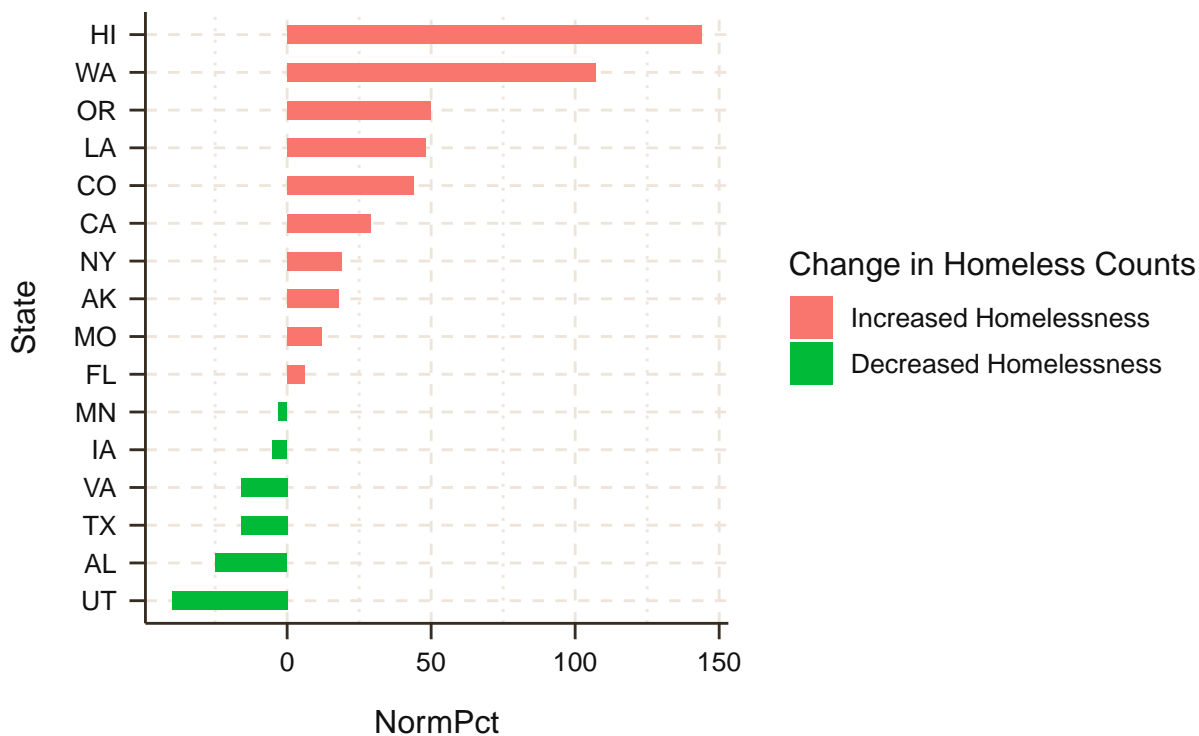
#ch_sub <- ch_sub %>%
# filter(((NormPct > 10) & (NormPct < 150)) | (NormPct < -10))

ch_sub2 <- ch_sub %>%
  filter(State %in% c('WA', 'CA', 'NY', 'FL', 'TX', 'MN', 'UT', 'AL', 'VA', 'LA', 'OR', 'HI', 'OR', 'CO'))

#Diverging Barcharts
div_plot <- ggplot(ch_sub2, aes(x=State, y=NormPct, label=NormPct)) +
  geom_bar(stat='identity', aes(fill=NormDir), width=.5) +
  scale_fill_manual(name="Change in Homeless Counts",
                    labels= c("Increased Homelessness", "Decreased Homelessness"),
                    values = c("above"="#f8766d", "below"="#00ba38")) +
  labs(title= "Chronic Homelessness: Increase vs. Decrease across States") +
  coord_flip()

print(div_plot)
```

## Chronic Homelessness: Increase vs. Decrease across States



```
div_plot <- ggplot(ch_sub2, aes(x=State, y=NormPct, label=NormPct)) +
  geom_bar(stat='identity', aes(fill=NormDir), width=.5) +
  scale_fill_manual(name="Change in Homeless Counts",
                    labels= c("Increased Homelessness", "Decreased Homelessness"),
                    values = c("above"="cadetblue", "below"="antiquewhite4")) +
  labs(title= "Chronic Homelessness: Increase vs. Decrease across States") +
  coord_flip()

print(div_plot)
```

## Chronic Homelessness: Increase vs. Decrease across States

