# miniproject2

*Shinhae Park*

*5/22/2019*

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```
library(tidyr)
library(readr)
library(forcats)
```

```
setwd("/cloud/project/Day3/MiniProject02-Healthcare/1_Dataset")
coverage<-read_csv("hcare_cov.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Location = col_character(),
##   `2013__Other Public` = col_character(),
##   `2014__Other Public` = col_character(),
##   `2015__Other Public` = col_character(),
##   `2016__Other Public` = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
spend<-read_csv("hcare_spend.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Location = col_character()
## )
## See spec(...) for full column specifications.
```

Let us check the data and the definitions of each variable.

```r
head(coverage) #US total excludes Puerto Rico
```

```
## # A tibble: 6 x 29
##   Location `2013__Employer` `2013__Non-Grou~ `2013__Medicaid`
##   <chr>               <dbl>            <dbl>            <dbl>
## 1 United ~        155696900         13816000         54919100
## 2 Alabama           2126500           174200           869700
## 3 Alaska             364900            24000            95000
## 4 Arizona           2883800           170800          1346100
## 5 Arkansas          1128800           155600           600800
## 6 Califor~         17747300          1986400          8344800
## # ... with 25 more variables: `2013__Medicare` <dbl>, `2013__Other
## #   Public` <chr>, `2013__Uninsured` <dbl>, `2013__Total` <dbl>,
## #   `2014__Employer` <dbl>, `2014__Non-Group` <dbl>,
## #   `2014__Medicaid` <dbl>, `2014__Medicare` <dbl>, `2014__Other
## #   Public` <chr>, `2014__Uninsured` <dbl>, `2014__Total` <dbl>,
## #   `2015__Employer` <dbl>, `2015__Non-Group` <dbl>,
## #   `2015__Medicaid` <dbl>, `2015__Medicare` <dbl>, `2015__Other
## #   Public` <chr>, `2015__Uninsured` <dbl>, `2015__Total` <dbl>,
## #   `2016__Employer` <dbl>, `2016__Non-Group` <dbl>,
## #   `2016__Medicaid` <dbl>, `2016__Medicare` <dbl>, `2016__Other
## #   Public` <chr>, `2016__Uninsured` <dbl>, `2016__Total` <dbl>
```

## Medicaid:

Includes those covered by Medicaid, Medical Assistance, Children's Health Insurance Plan (CHIP) or any kind of government-assistance plan for those with low incomes or a disability, as well as those who have both Medicaid and another type of coverage, such as dual eligibles who are also covered by Medicare.

## Medicare:

Includes those covered by Medicare, Medicare Advantage, and those who have Medicare and another type of non-Medicaid coverage where Medicare appears to be the primary payer. Excludes seniors who also report employer-sponsored coverage and full-time work, and those covered by Medicare and Medicaid (dual eligibles).

## Employer:

Includes those covered through a current or former employer or union, either as policyholder or as dependent

## Other Public:

Includes those covered under the military or Veterans Administration

## Non-Group:

Includes those covered by a policy purchased directly from an insurance company, etiher as policyholder or as dependent

## Uninsured:

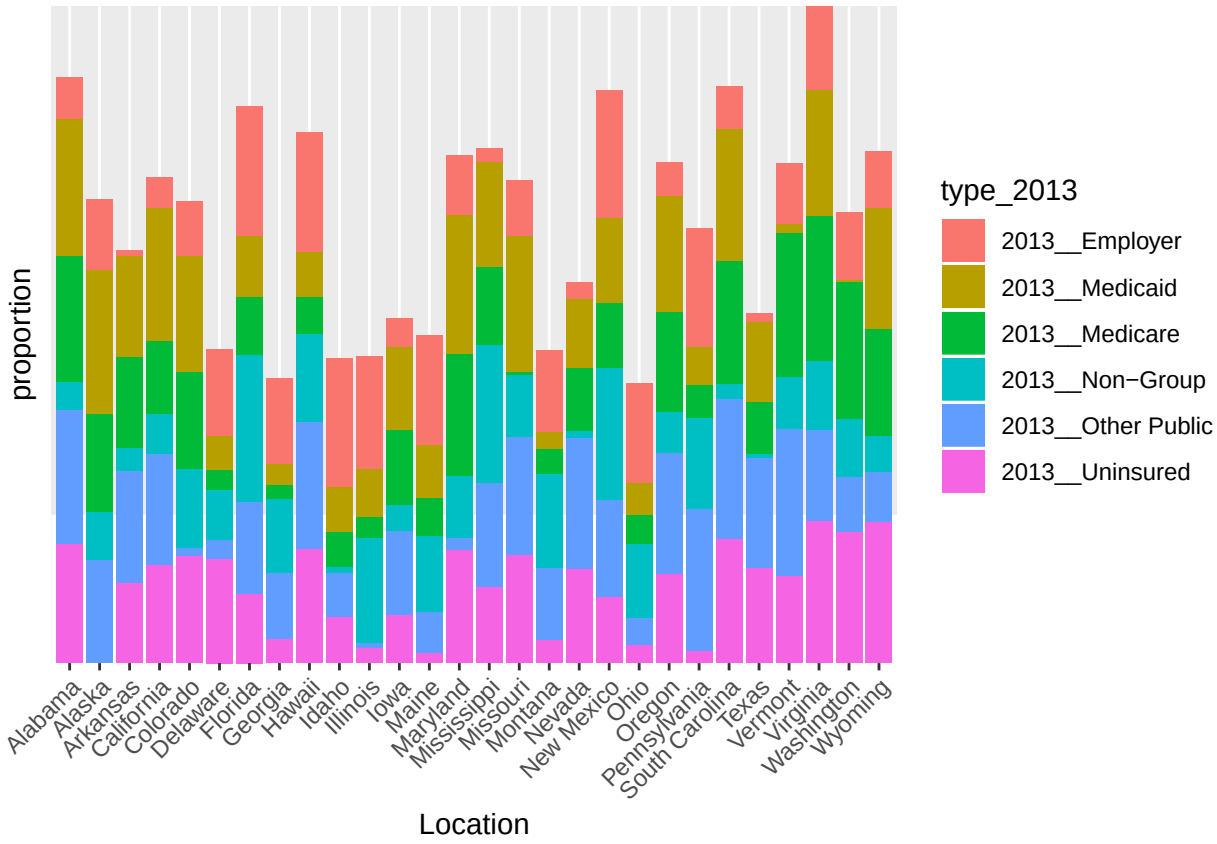
Includes those without health insurance and those who have coverage under the Indian Health Service only.

## N/A: Estimates with relative standard errors greater than 30% are not provided.

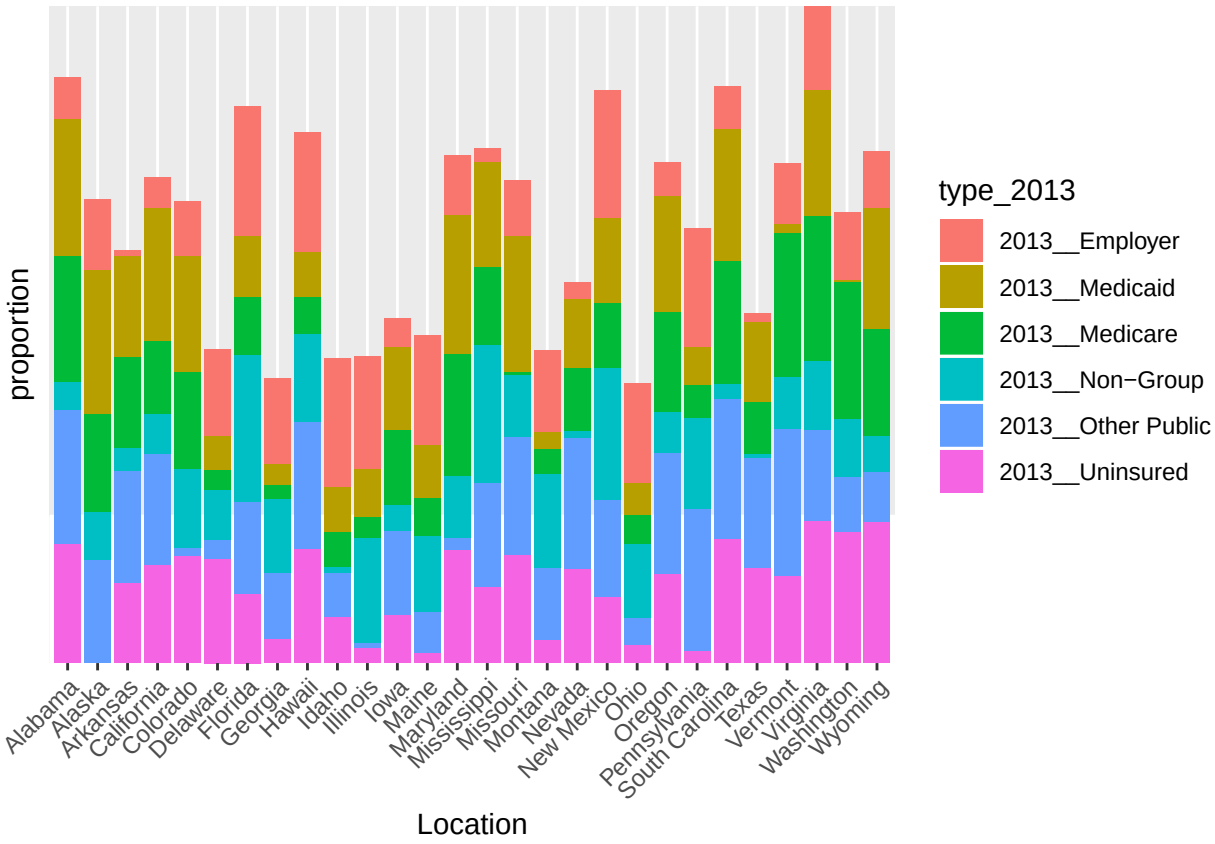
```
#remove N/A

coverage2=coverage[-which(coverage=="N/A",arr.ind=TRUE)[,1],]
```

```
coverage2 %>%
  select(1:7) %>%
  filter(Location!="United States") %>%
  gather("type_2013","amount", -Location) %>%
  group_by(type_2013) %>%
  ggplot(aes(x=Location,y=amount))+
  geom_bar(aes(fill=type_2013),stat="identity", position = "stack")+
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=1))+
  theme(axis.text.y=element_blank(),axis.ticks.y=element_blank())+
  ylab("proportion")
```




```
coverage2 %>%
  select(1:7) %>%
  filter(Location!="United States") %>%
  gather("type_2013","amount", -Location) %>%
  group_by(type_2013) %>%
  ggplot(aes(x=Location,y=amount))+
  geom_col(aes(fill=type_2013))+
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=1))+
  theme(axis.text.y=element_blank(),axis.ticks.y=element_blank())+
  ylab("proportion")
```




Now, I want to see how the coverage changed from 2013 to 2016 in Illinois.

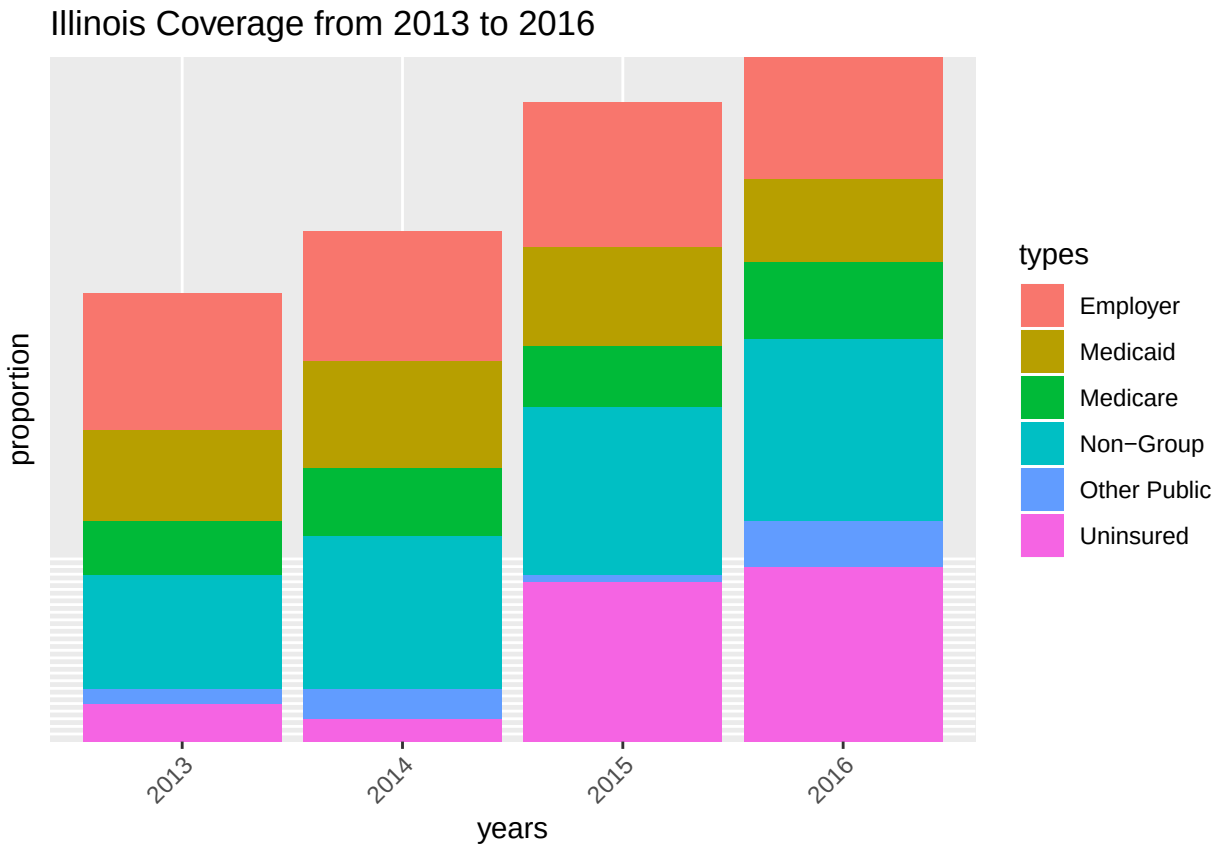
```
cov_illinois<-coverage2 %>%
  filter(Location=="Illinois") %>%
  gather("types","amount", -Location) %>%
  filter(!grepl("Total",types)) %>%
  select(-c(Location))

#sp_illinois<-spend %>%
#  filter(Location=="Illinois")
```

```
cov_illinois[,"years"]<- c(rep("2013",6),rep("2014",6),rep("2015",6),rep("2016",6))
```

```
#which(grepl("Employer",cov_illinois$types))
cov_illinois$types[which(grepl("Employer",cov_illinois$types))]<-"Employer"
cov_illinois$types[which(grepl("Non-Group",cov_illinois$types))]<-"Non-Group"
cov_illinois$types[which(grepl("Medicaid",cov_illinois$types))]<-"Medicaid"
cov_illinois$types[which(grepl("Medicare",cov_illinois$types))]<-"Medicare"
cov_illinois$types[which(grepl("Other Public",cov_illinois$types))]<-"Other Public"
cov_illinois$types[which(grepl("Uninsured",cov_illinois$types))]<-"Uninsured"
```

```
cov_illinois %>%
  group_by(years) %>%
  ggplot(aes(x=years,y=amount))+
  geom_bar(aes(fill=types),stat="identity", position = "stack")+
  theme(axis.text.x=element_text(angle=45,hjust=1,vjust=1))+
  theme(axis.text.y=element_blank(),axis.ticks.y=element_blank())+
  ylab("proportion")+ggtitle("Illinois Coverage from 2013 to 2016")
```




**Question 1: Is there a relationship between healthcare coverage and healthcare spending in the United States?**

**Question 2: How does the spending distribution change across geographic regions in the United States?**

**Question 3: Does the relationship between healthcare coverage and healthcare spending in the United States change from 2013 to 2014?**

```
spend2013_14<-spend %>%
  select(Location,`2013__Total Health Spending`, `2014__Total Health Spending`) %>% filter(Location=="U
```

```
data2013_14<-coverage %>%
  select(Location,`2013__Total`, `2014__Total`) %>%
  filter(Location=="United States") %>%
  full_join(spend2013_14, by="Location") %>%
  ggplot(aes=("Location") )+geom_bar(stat="identity",fill="pink")
```