R Notebook on National Opioid Death Simple EDA

ISABEL METZGER

Making graphs used in glimpse of US opioid deaths/ state rates

```r
library(noncensus)
data("states")
overdosedeath <- death2015 %>% left_join(states, by="state")
fullopioiddeathset_2013_2015 <- overdosedeath %>% right_join(Xdeath_2013_2014)

library(ggthemes)
# re-order levels
reorder_size <- function(x) {
        factor(x, levels = names(sort(table(x), decreasing = TRUE)))
}

head(fullopioiddeathset_2013_2015,2)
```

```
## # A tibble: 2 x 18
##   state  Range.2015 Rate.2015 Deaths.2015           name  region
##   <chr>       <chr>     <dbl>       <dbl>          <chr>  <fctr>
## 1    ND 2.8 to 11.0       8.6          61 North Dakota Midwest
## 2    NE 2.8 to 11.0       6.9         126      Nebraska Midwest
## # ... with 12 more variables: division <fctr>, capital <chr>, area <chr>,
## #   population <chr>, Rate.2014 <dbl>, Deaths.2014 <dbl>,
## #   Range.2014 <chr>, Rate.2013 <dbl>, Deaths.2013 <dbl>,
## #   Range.2013 <chr>, Change <dbl>, Significant <chr>
```

```r
longset <- fullopioiddeathset_2013_2015 %>% gather(Deaths.2013,Deaths.2014,Deaths.2015, key="Year", valu

mycols3
```

```
## [1] "#c6d4e1" "#2f2016" "#fcfaea" "#456789"
```

```r
cols <-c("#9D1F2F","#d6604d", "#92c5de","#0571b0")
longset$year <- longset$Year
longset$Year <- gsub("Deaths.2013", 2013, longset$Year)
longset$Year <- gsub("Deaths.2014", 2014, longset$Year)
longset$Year <- gsub("Deaths.2015", 2015, longset$Year)
 res <- longset %>% dplyr::select(Year, Deaths, region)
 res$region <- as.character(res$region)
 res$Year <- as.numeric(res$Year)
  colnames(res) <- c("Year", "Opioid.Deaths", "Region")
 res %>% head(5)
```

```
## # A tibble: 5 x 3
##    Year Opioid.Deaths  Region
##   <dbl>         <dbl>   <chr>
## 1  2013            20 Midwest
## 2  2013           117 Midwest
## 3  2013            55 Midwest
## 4  2013           275 Midwest
## 5  2013          2446   South
```

```
# par(mar=c(5.1, 4.1, 4.1, 7.1), xpd=TRUE)
# barplot(prop, col=heat.colors(length(rownames(prop))), width=2, beside=TRUE)
# legend("topright",inset=c(-0.25,0), fill=heat.colors(length(rownames(prop))), legend=rownames(data))

# longset %>% ggplot(aes(region, Deaths)) + geom_point(aes(col=state)) + facet_wrap(~Year)
longset %>% filter(Year==2013) %>% mutate(state = state, Deaths)%>% ggplot(aes(state,Deaths)) + geom_col
        theme(axis.text = element_text(hjust = 1, size=6.7, angle=20)) + scale_fill_manual(values=cols)
```
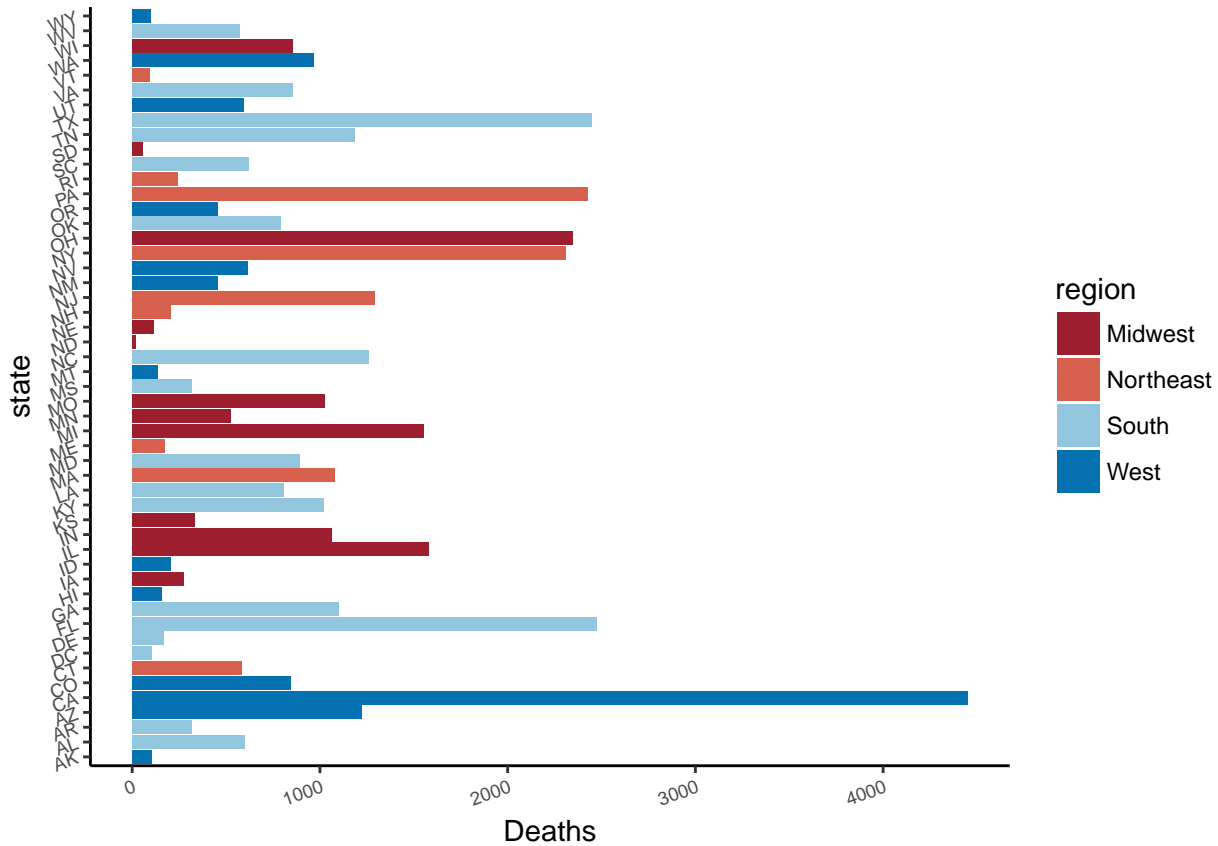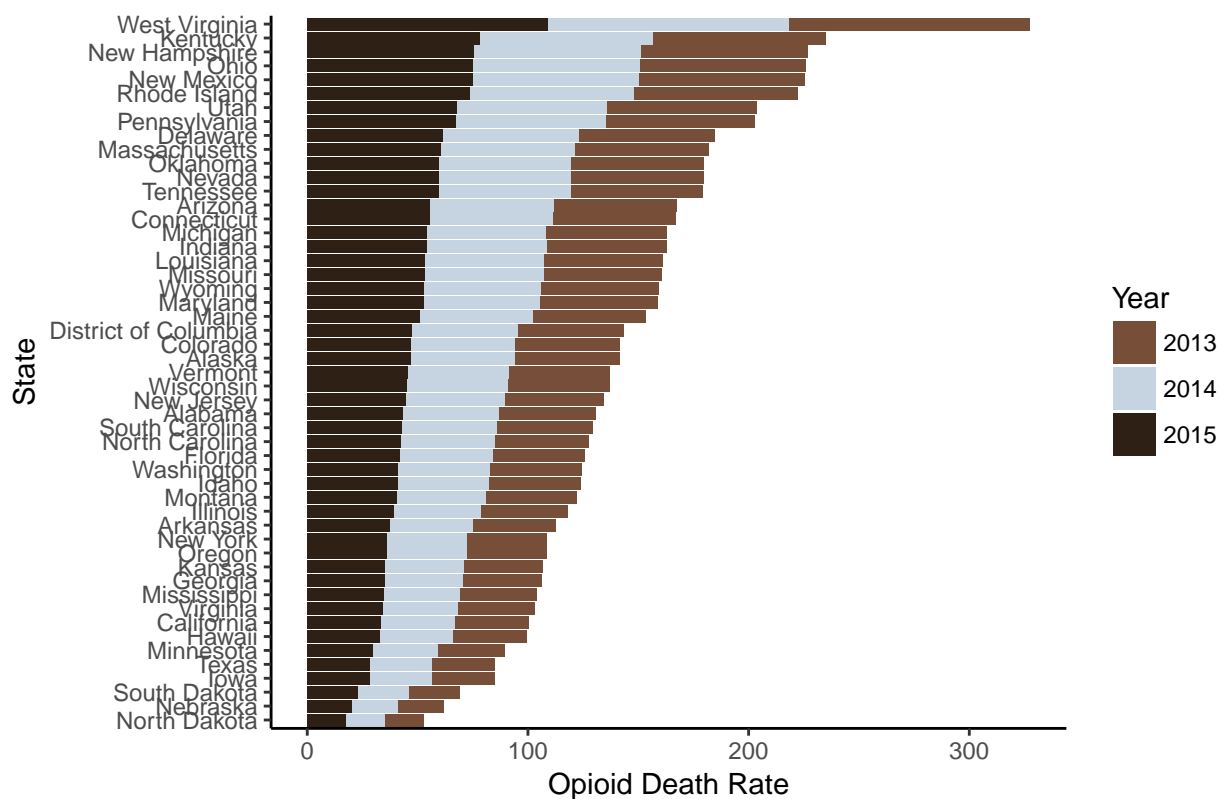


```
longset %>% filter(Year==2014) %>% mutate(state = reorder(state, Deaths))%>% ggplot(aes(state, Deaths))
        theme(axis.text = element_text(hjust = 1, size=6.7, angle=20)) + scale_fill_manual(values=cols)
```
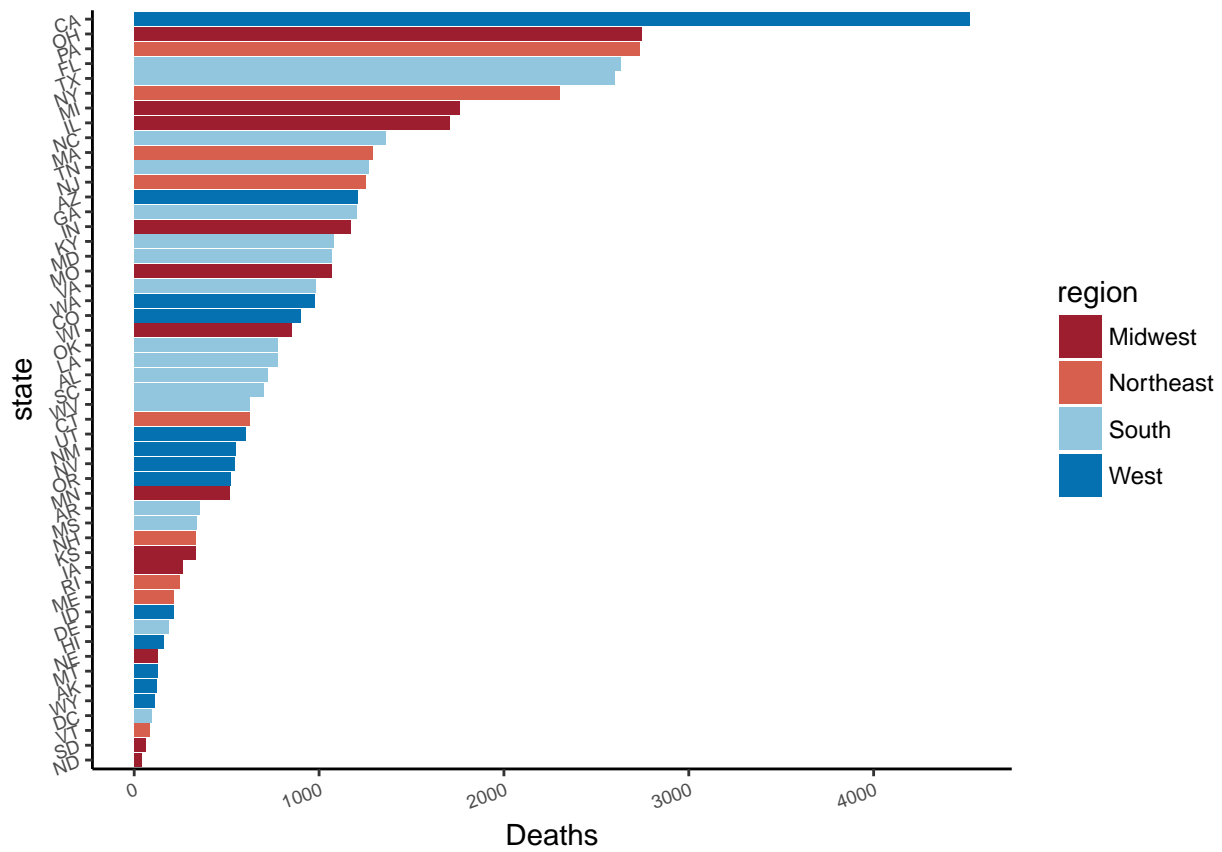
```
longset %>% filter(Year==2015) %>% mutate(state = reorder(state, Deaths))%>% ggplot(aes(state, Deaths))
        theme(axis.text = element_text(hjust = 1, size=6.7, angle=20)) + scale_fill_manual(values=cols)
```

```
longset %>% gather(Rate.2013,Rate.2014,Rate.2015, key="yr", value="Rate") %>% mutate(name=reorder(name,
```
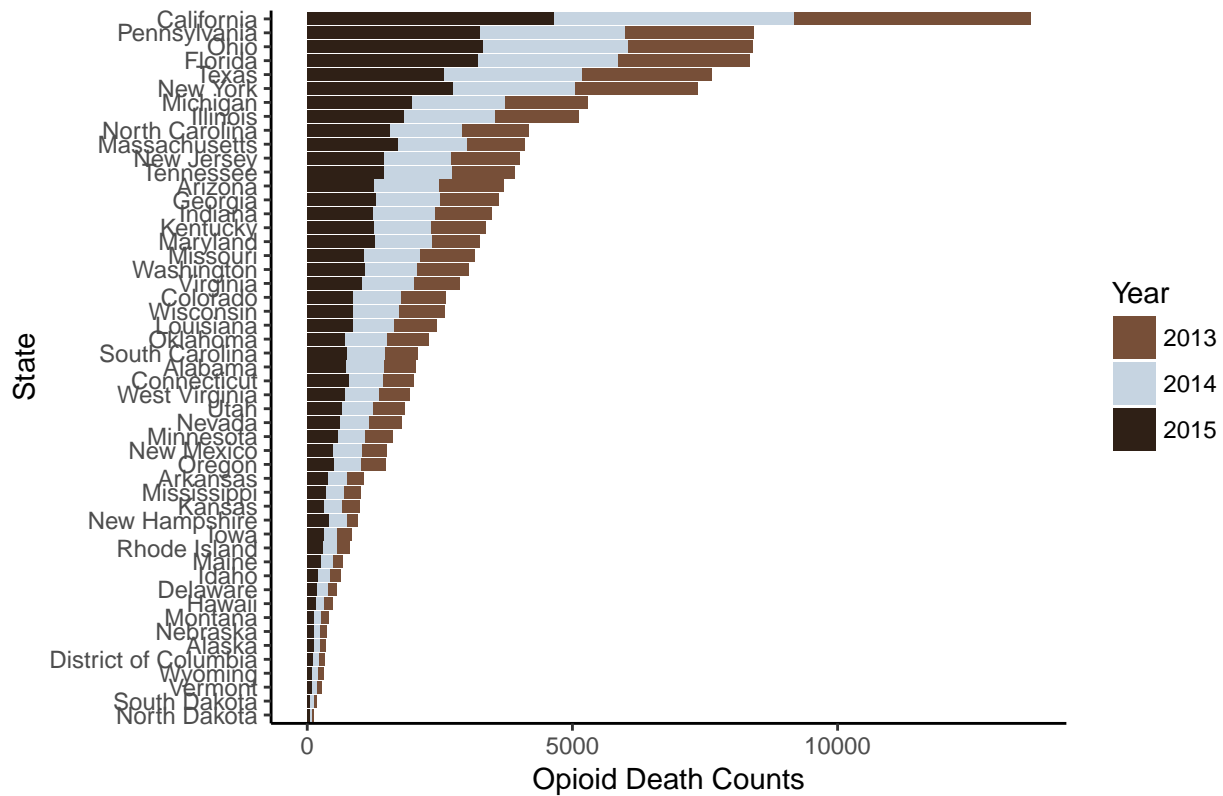
Age and population adjusted Opioid Death Rates by State
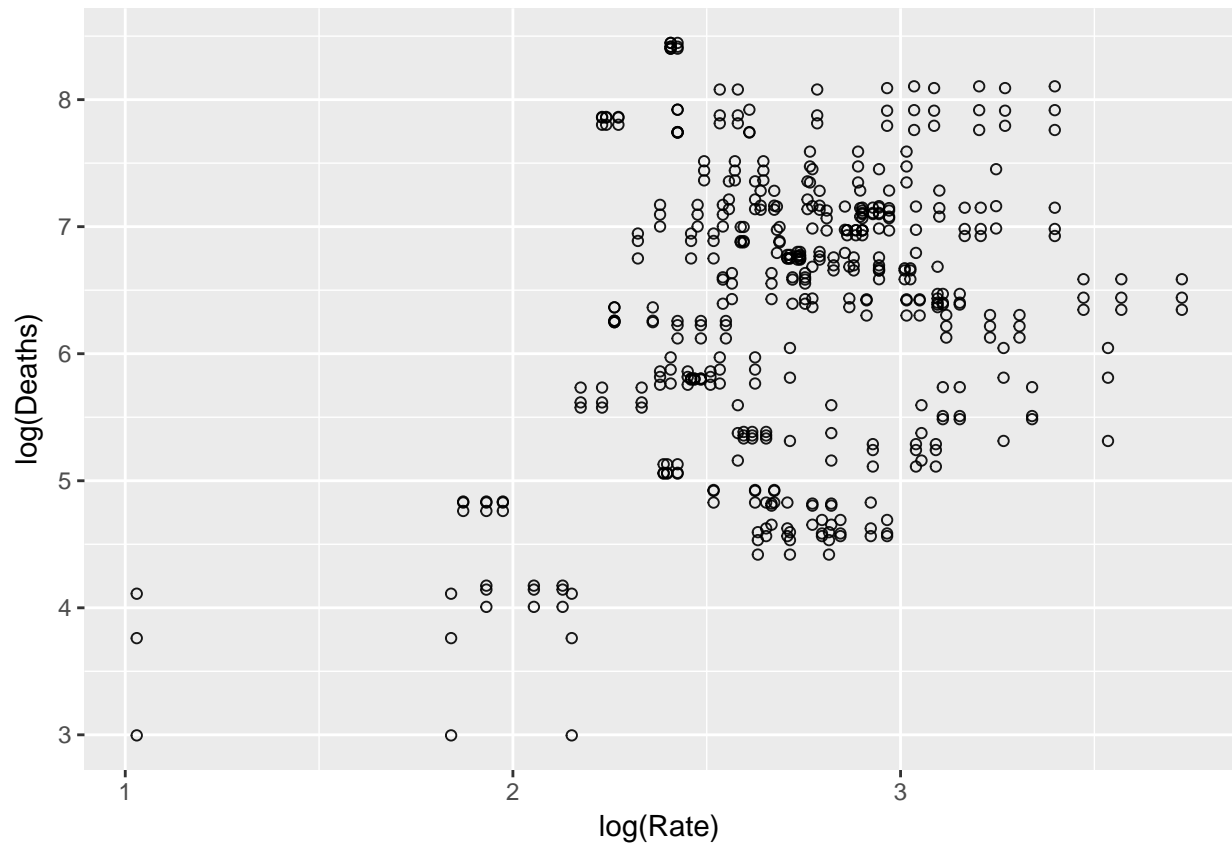


```
longset  %>% mutate(name=reorder(name, Deaths)) %>% ggplot(aes(name, Deaths, fill=Year)) + geom_col() +
```
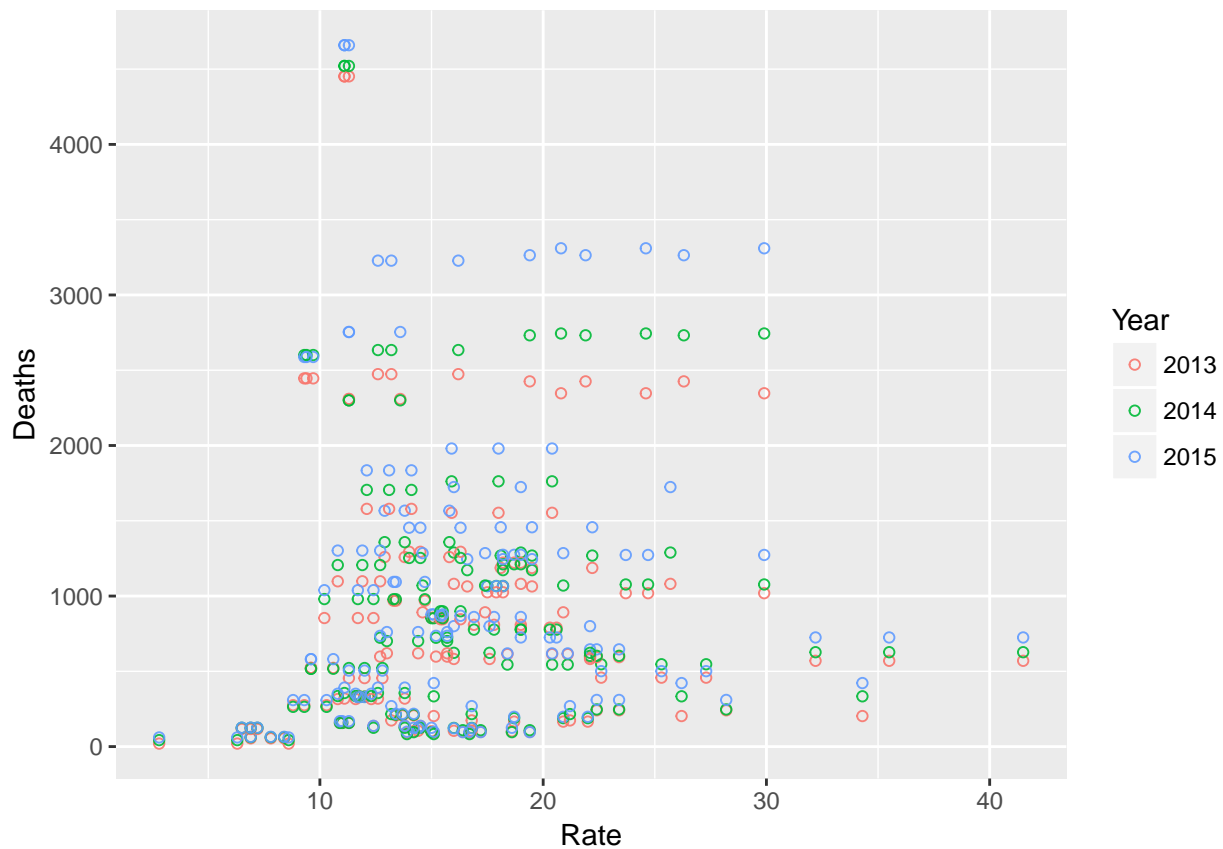
Opioid Death Raw Counts by State

```
longset %>% gather(Rate.2013,Rate.2014,Rate.2015, key="yr", value="Rate") %>% ggplot(aes(log(Rate), log
```

```
longset %>% gather(Rate.2013,Rate.2014,Rate.2015, key="yr", value="Rate") %>% ggplot(aes(Rate, Deaths,
```

```r
rates1 <- longset %>% gather(Rate.2013,Rate.2014,Rate.2015, key="yr", value="Rate")
summary(rates1$Rate)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.80   12.10   15.20   16.25   19.00   41.50
```

```r
library(readr)
stategrid <- read.csv("state-grid-coordinates.tsv", stringsAsFactors = FALSE, sep="\t")
head(stategrid)
```

```
##   state x y
## 1    AK 1 1
## 2    WA 1 3
## 3    OR 1 4
## 4    CA 1 5
## 5    HI 1 8
## 6    ID 2 3
```

```r
plot(stategrid$x, stategrid$y, type="n")
text(stategrid$x, stategrid$y, stategrid$state)
```
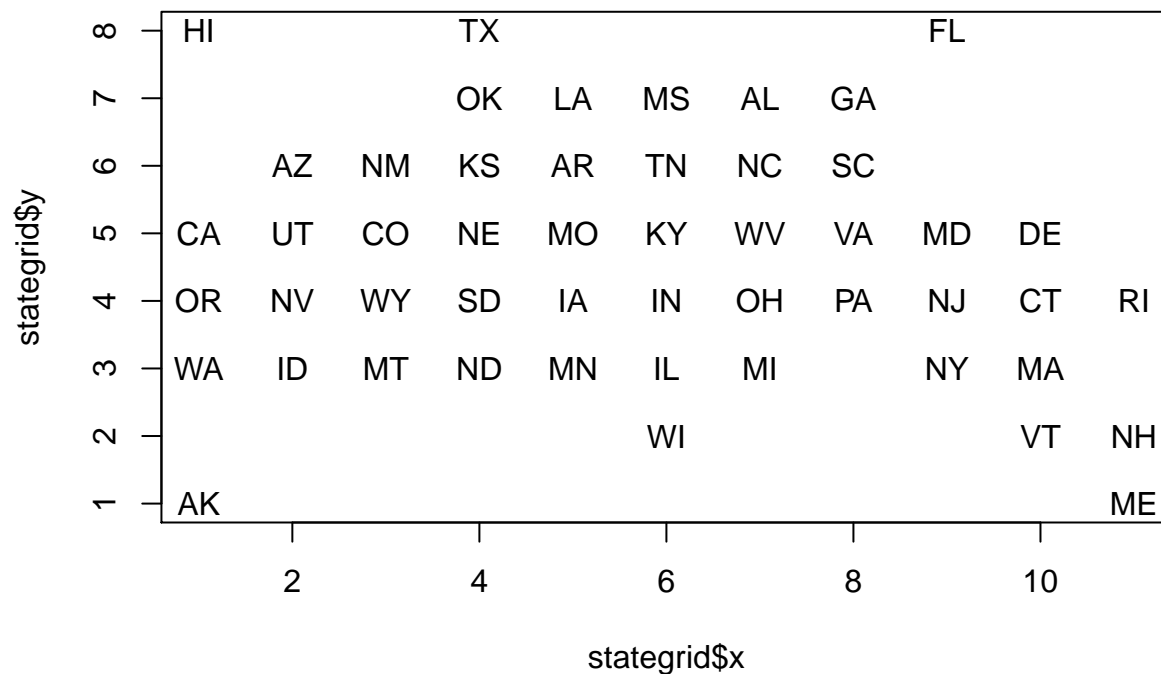
```
stategrid$ysideup <- 12 - stategrid$y
plot(stategrid$x, stategrid$ysideup, type="n")
text(stategrid$x, stategrid$ysideup, stategrid$state)
```
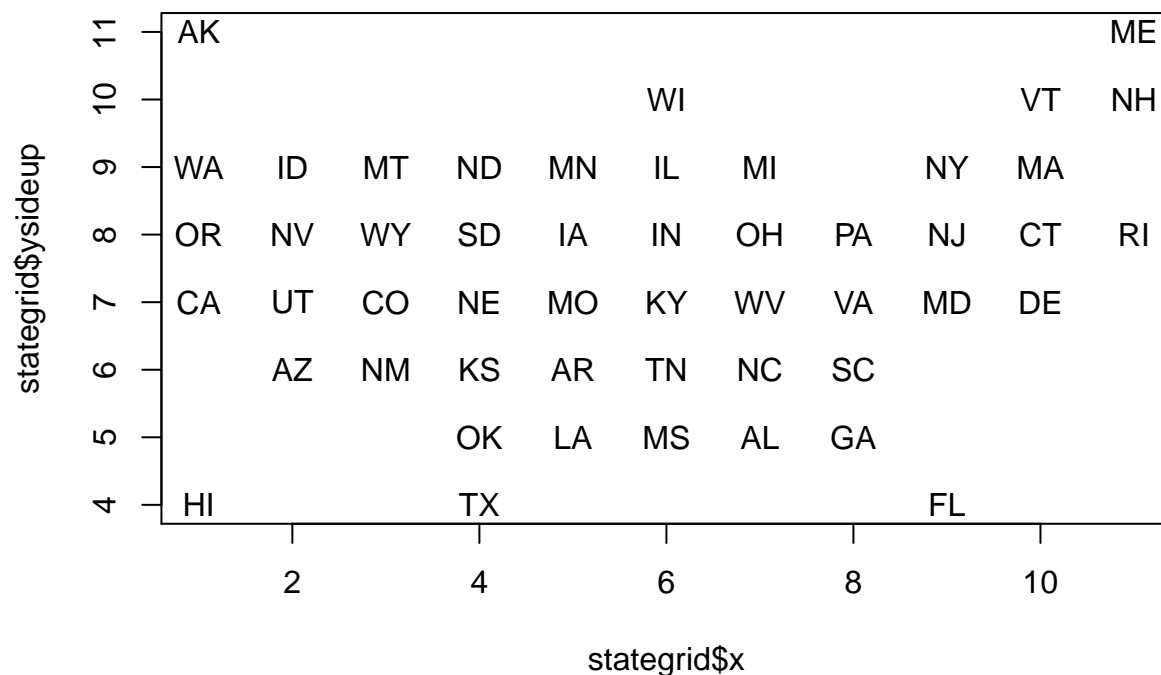


```
symbols(stategrid$x, stategrid$ysideup, squares = rep(1, dim(stategrid)[1]), inches=FALSE, asp=1, bty="
text(stategrid$x, stategrid$ysideup, stategrid$state)
```

```
AK                              ME
              WI            VT NH
WA ID MT ND MN IL MI      NY MA
OR NV WY SD IA IN OH PA NJ CT RI
CA UT CO NE MO KY WV VA MD DE
   AZ NM KS AR TN NC SC
      OK LA MS AL GA
HI       TX            FL
```

```r
head(fullopioiddeathset_2013_2015)
```

```
## # A tibble: 6 x 18
##   state  Range.2015 Rate.2015 Deaths.2015          name  region
##   <chr>       <chr>     <dbl>       <dbl>         <chr>  <fctr>
## 1    ND 2.8 to 11.0       8.6          61 North Dakota Midwest
## 2    NE 2.8 to 11.0       6.9         126      Nebraska Midwest
## 3    SD 2.8 to 11.0       8.4          65 South Dakota Midwest
## 4    IA 2.8 to 11.0      10.3         309          Iowa Midwest
## 5    TX 2.8 to 11.0       9.4        2588         Texas   South
## 6    MN 2.8 to 11.0      10.6         581     Minnesota Midwest
## # ... with 12 more variables: division <fctr>, capital <chr>, area <chr>,
## #   population <chr>, Rate.2014 <dbl>, Deaths.2014 <dbl>,
## #   Range.2014 <chr>, Rate.2013 <dbl>, Deaths.2013 <dbl>,
## #   Range.2013 <chr>, Change <dbl>, Significant <chr>
```
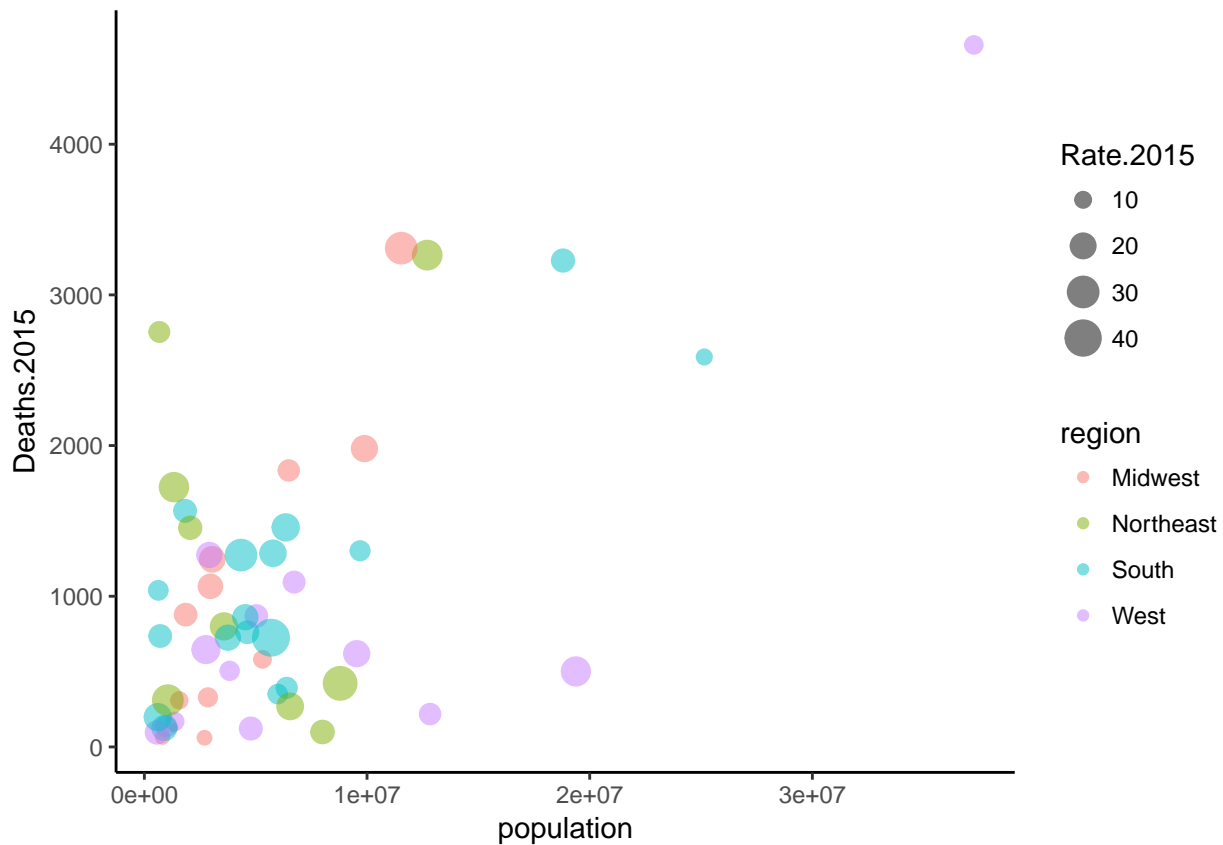
```r
fullopioiddeathset_2013_2015$population <- as.numeric(fullopioiddeathset_2013_2015$population)
```

```r
fullopioiddeathset_2013_2015%>% ggplot(aes(population, Deaths.2015)) + geom_point(aes(size=Rate.2015, co
```

```
longset$popoverarea <- as.numeric(longset$population)/as.numeric(longset$area)
summary(100*(longset$popoverarea))
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.      Max.
##     575.6    3468.2   10742.3   43062.1   23475.5 1314691.1
```

```
longset$highdense <- longset$popoverarea
longset$highdense <- ifelse(longset$popoverarea > 230, "Dense","NotDense")
fullopioiddeathset_2013_2015$Change.2014.2015 <-
  100*((fullopioiddeathset_2013_2015$Rate.2015/fullopioiddeathset_2013_2015$Rate.2014)-1)

table(fullopioiddeathset_2013_2015$Significant)
```

```
##
## Not Significant      Significant
##             37               14
```

```
longset$population <- as.numeric(longset$population)
DF <- fullopioiddeathset_2013_2015 %>% right_join(longset,by = c("state", "Range.2015", "Rate.2015", "na

interested.names <- c("Change.2014.2015","Significant","Change", "highdense", "name", "region", "Deaths
finalDF <- unique(DF[interested.names])
sample( c(1:10) , 51 , replace=T)
```

```
##  [1]  1  8  8  5  8  5  6 10 10  9  4  1  2  8  9  6  8  1  7  2  2  4  7
## [24]  2 10  4  4  3 10  9  4  6  9  7  7  1  2  2  9 10  2  6  1  3  2  7
## [47]  9  3  9  2  2
```

```r
fullopioiddeathset_2013_2015$Change.2014.2015 <-
  100*((fullopioiddeathset_2013_2015$Rate.2015/fullopioiddeathset_2013_2015$Rate.2014)-1)
fullopioiddeathset_2013_2015$Change.2013.2014 <-
  100*((fullopioiddeathset_2013_2015$Rate.2014/fullopioiddeathset_2013_2015$Rate.2013)-1)

dffff <- fullopioiddeathset_2013_2015 %>% gather(Change.2013.2014, Change.2014.2015, key="YearChange",
length(dffff$YearChange)
```

```
## [1] 102
```

```r
# dffff %>% select(YearChange, Percentage)

n <- 5
a<- rep(1:9, each=n)
b <- rep(2013:2014, each=51)

dffff$toadd<- c(a,a,1:9, 1, 2,3)
# c <- rep(1:5, each=n)
# d <- rep(1:5, each=n)
# dffff$toadd <- c(a, a, a, a, 5, 6)
length(dffff$toadd)
```

```
## [1] 102
```

```r
dffff$numbers <- paste(b, dffff$toadd)
dffff$numbers <- gsub(" ", ".", dffff$numbers)
dffff$numbers <- as.numeric(dffff$numbers)


DFDF <- dffff[c("numbers","Percentage")]
```

```r
# dffff$year=sample( c(13:15) , 51 , replace=T)
#  changes2=data.frame(name=finalDF$name, Region=finalDF$region, Change=finalDF$Change.,Year=sample( c(
# # Plotly library
# library(plotly)
Deaths <- c(fullopioiddeathset_2013_2015$Deaths.2013, fullopioiddeathset_2013_2015$Deaths.2015)
# # Make the graph
# my_graph=plot_ly(x=b, y=a , mode="markers" , size=abs(a), color=ifelse(a>0,"red","green") )
#
#Change hover mode in the layout argument :
# newdf <- rbind(changes2, changes)
# newdf$Deaths <- Deaths
# summary(newdf$Change)
DFDF %>% ggplot(aes(numbers, Percentage)) + geom_point(aes(size=abs((dffff$Percentage -.0001)*2901), col
```

% Change in Opioid Death Rates
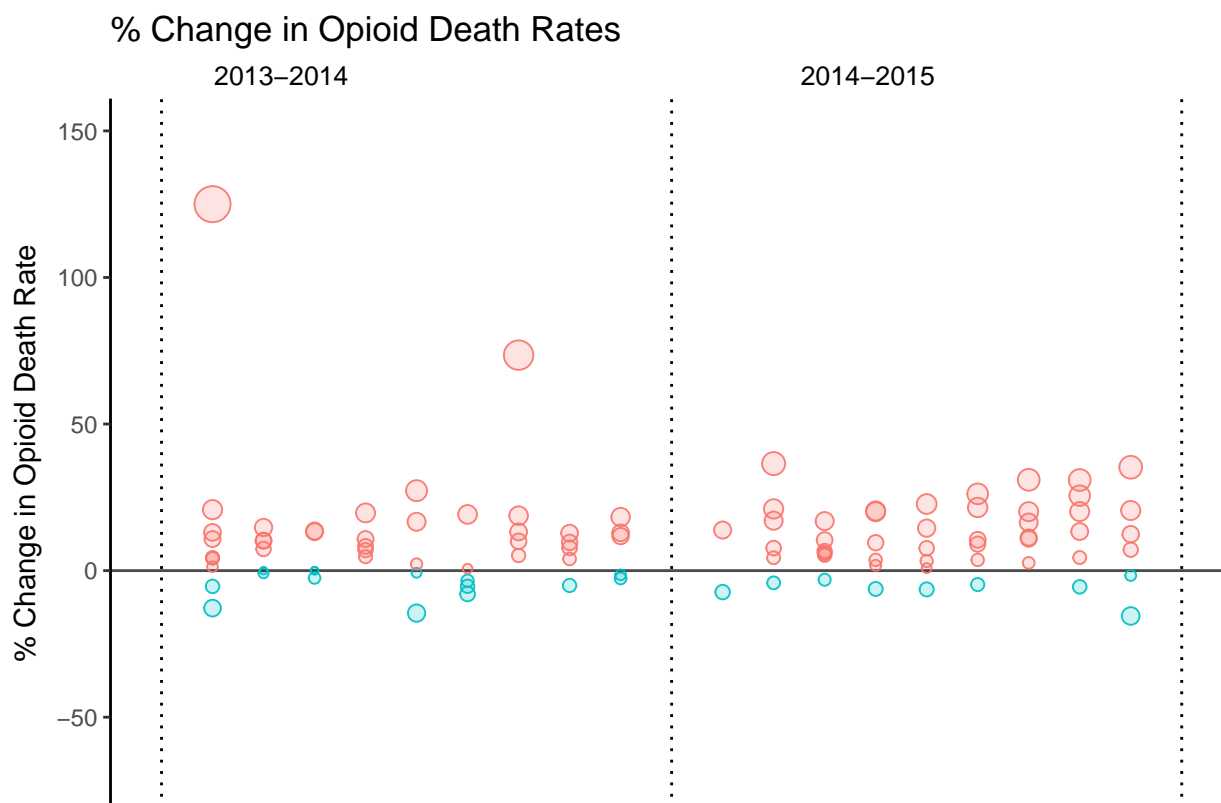

```
my_mean=aggregate(longset$Deaths , by=list(longset$region) , mean) ; colnames(my_mean)=c("region" , "mea
my_sd=aggregate(longset$Deaths  , by=list(longset$region) , sd) ; colnames(my_sd)=c("region" , "sd")
my_info=merge(my_mean , my_sd , by.x=1 , by.y=1)

# Make the plot
ggplot(longset) +
        geom_point(aes(x = region, y = Deaths, size=Deaths) , colour=rgb(0,1,0.60,0.34), alpha=0.2) +
        geom_point(data = my_info, aes(x=region , y = mean) , colour = rgb(0.7,0.6,0.9,0.7) , size = 8)
        geom_errorbar(data = my_info, aes(x = region, y = sd, ymin = mean - sd, ymax = mean + sd), colou
```
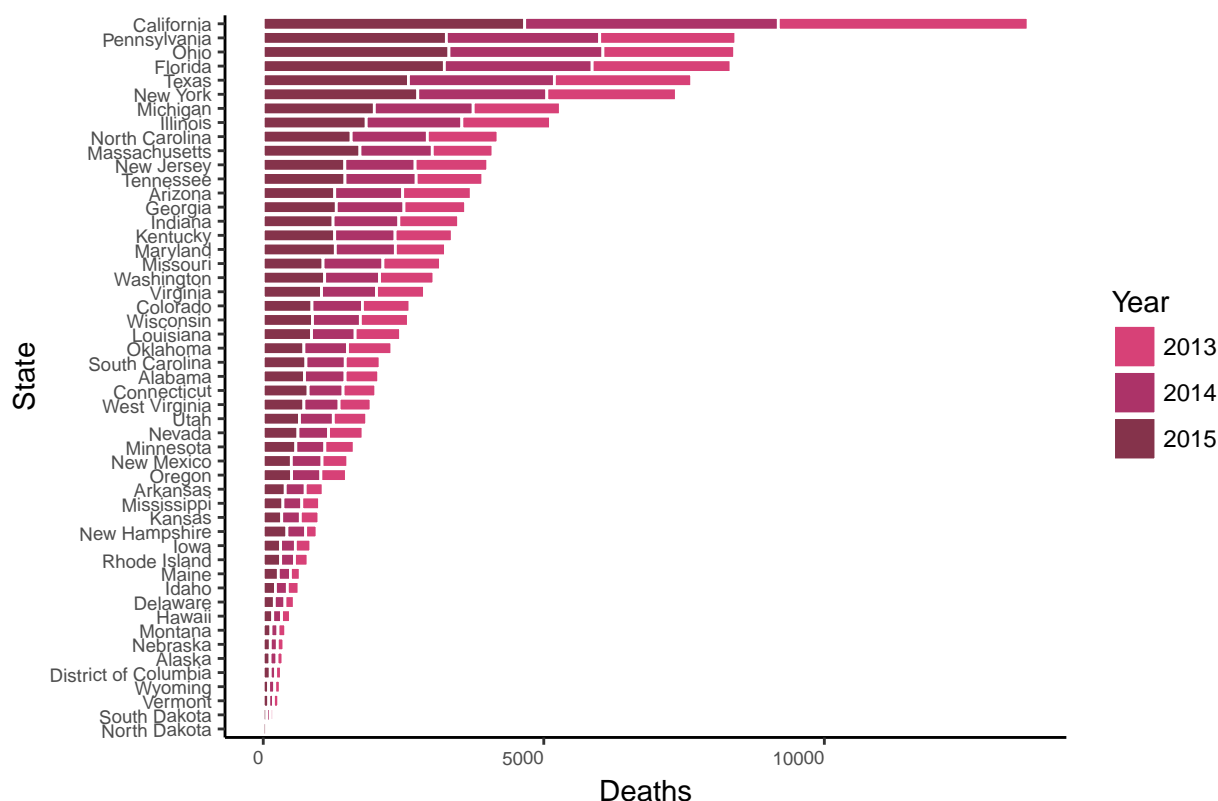
Range of Opioid.Deaths by US Demographic Region, 2013–2015
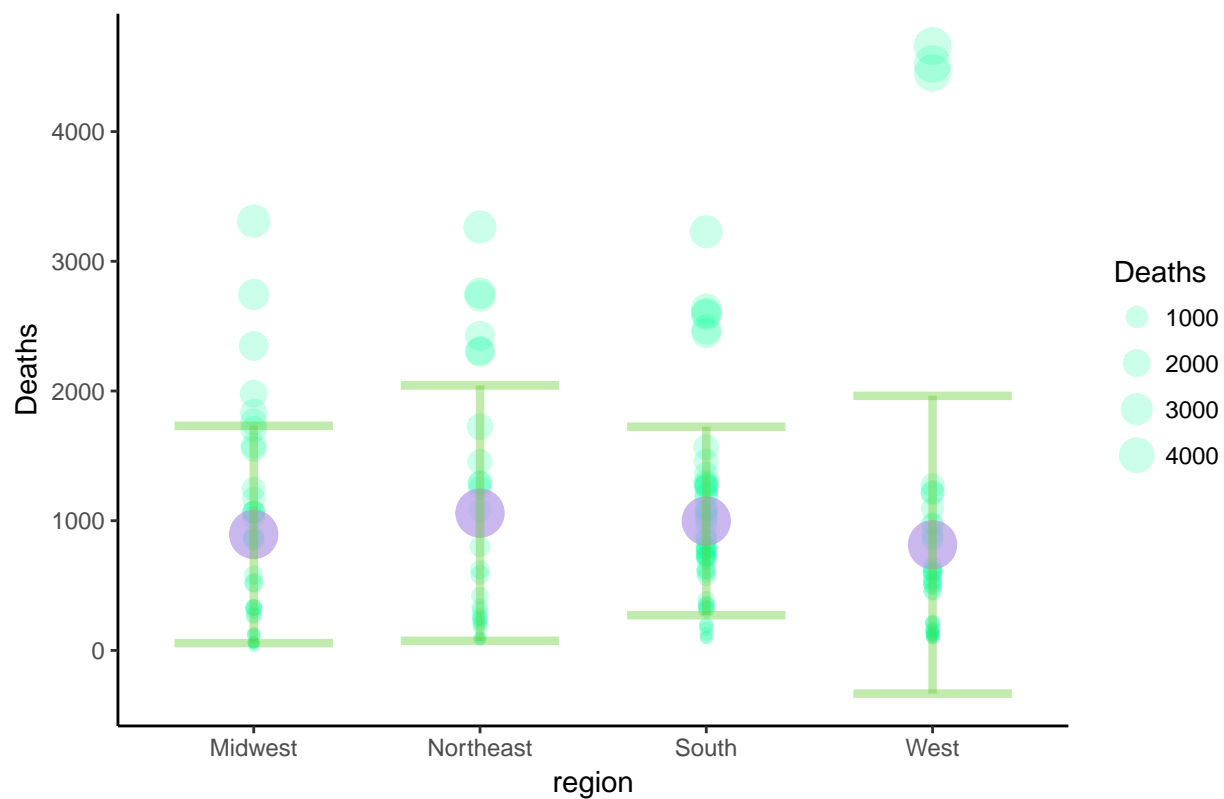

```
df <- longset %>% dplyr::select(name, Year, region, Deaths)
unique(df) %>% ggplot(aes(x=reorder(name,Deaths),y=Deaths, fill=Year)) + geom_col(col="white", alpha=0.8
        theme(axis.text = element_text(hjust = 1, size=7, angle=1)) + scale_fill_manual(values=c("#ce12
```

Opioid Deaths 2013 – 2015


```r
overdosegrid <- fullopioiddeathset_2013_2015 %>% merge(stategrid, by="state")
symbols(overdosegrid$x, overdosegrid$ysideup,
        squares = rep(1, dim(overdosegrid)[1]),
        inches=FALSE,
        asp=1,
        bty="n",
        xaxt="n", yaxt="n",
        xlab="", ylab="",
        bg=overdosegrid$col,
        fg="#ffffff")
labeltext <- paste(overdosegrid$state, "\n", format(overdosegrid$Rate.2015, 2), sep="")
text(overdosegrid$x, overdosegrid$ysideup, labeltext, cex=.6, col="black")
```

AK									ME	
16.0									21.2	
					WI			VT	NH	
					15.5			16.7	34.3	
WA	ID	MT	ND	MN	IL	MI		NY	MA	
14.7	14.2	13.8	8.6	10.6	14.1	20.4		13.6	25.7	
OR	NV	WY	SD	IA	IN	OH	PA	NJ	CT	RI
12.0	20.4	16.4	8.4	10.3	19.5	29.9	26.3	16.3	22.1	28.2
CA	UT	CO	NE	MO	KY	WV	VA	MD	DE	
11.3	23.4	15.4	6.9	17.9	29.9	41.5	12.4	20.9	22.0	
	AZ	NM	KS	AR	TN	NC	SC			
	19.0	25.3	11.8	13.8	22.2	15.8	15.7			
			OK	LA	MS	AL	GA			
			19.0	19.0	12.3	15.7	12.7			
HI			TX				FL			
11.3			9.4				16.2			

```
summary(overdosegrid$Rate.2015)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.90   12.93   16.10   17.80   21.12   41.50
```
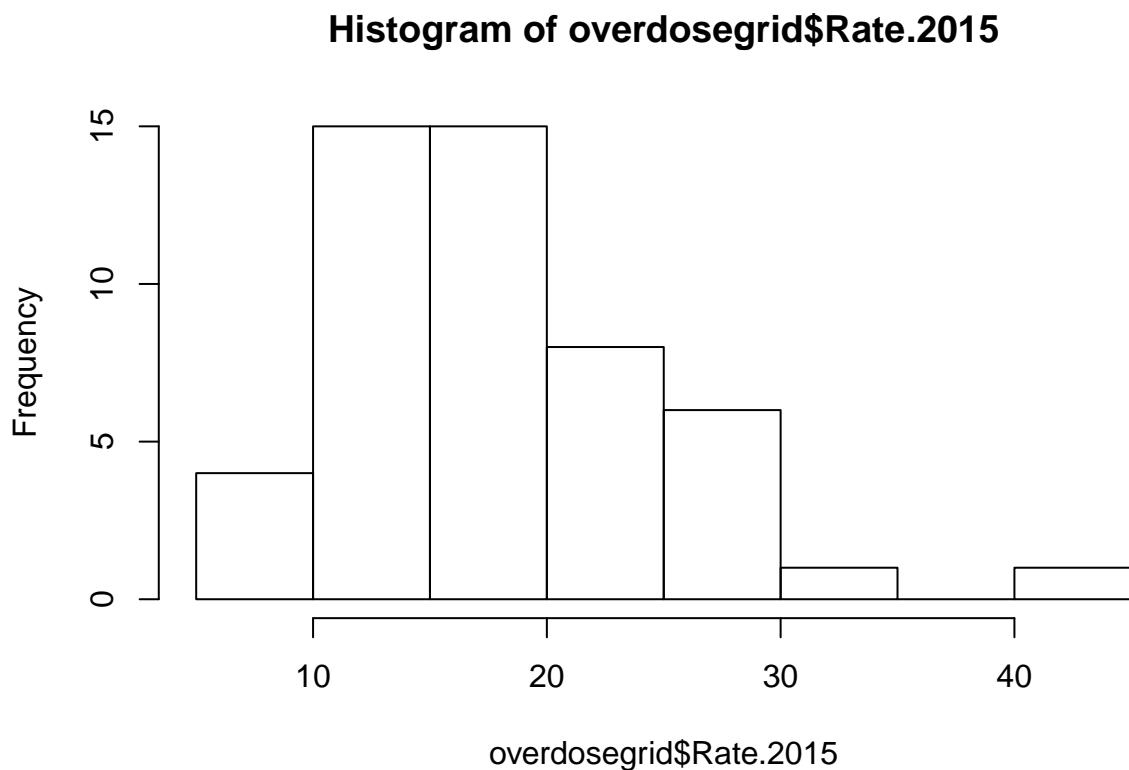
```
summary(overdosegrid$Rate.2014)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.30   12.03   15.15   16.10   18.85   35.50
```

```
summary(overdosegrid$Rate.2013)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.80   11.30   14.50   14.87   17.73   32.20
```

```
hist(overdosegrid$Rate.2015)
```

Histogram of overdosegrid$Rate.2015
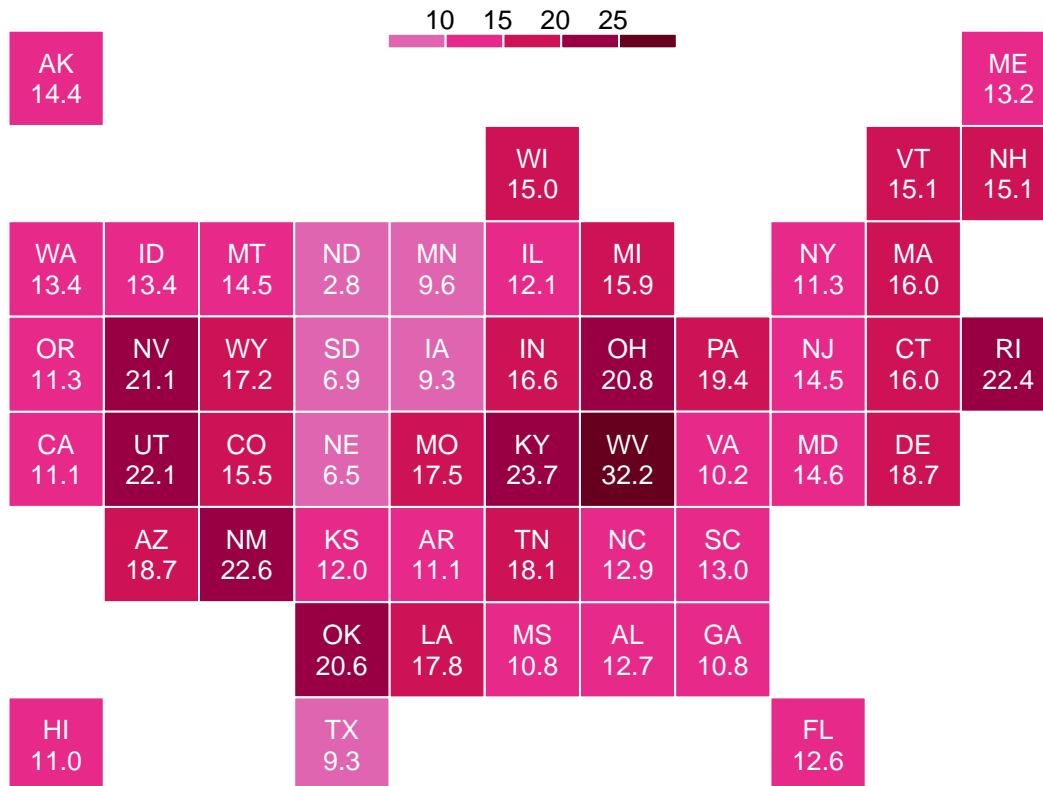


```
overdosegrid$col <- sapply(overdosegrid$Rate.2013, function(x) {
    # if (x < 5)
    #    {
    #      col <- "#d4b9da"
    # } else
      if (x < 10) {
    #      col <- "#c994c7"
    # } else if (x < 10) {
        col <- "#df65b0"
    } else if (x < 15) {
        col <- "#e7298a"
    } else if (x < 20) {
        col <- "#ce1256"
    } else if (x < 25) {
```

```r
        col <- "#980043"
    } else {
        col <- "#67001f"
    }
    return(col)
})
# Start layout.
par(mar=c(0,0,0,0), bg="white")
plot(0:1, 0:1, type="n", xlab="", ylab="", axes=FALSE, asp=1)
# Draw map like before.
par(new=TRUE, plt=c(0, 1, 0, 1))
symbols(overdosegrid$x, overdosegrid$ysideup,
        squares = rep(1, dim(overdosegrid)[1]),
        inches=FALSE,
        asp=1,
        bty="n",
        xaxt="n", yaxt="n",
        xlab="", ylab="",
        bg=overdosegrid$col,
        fg="#ffffff")
labeltext <- paste(overdosegrid$state, "\n", format(overdosegrid$Rate.2013, 2), sep="")
text(overdosegrid$x, overdosegrid$ysideup, labeltext, cex=.8, col="#ffffff")
# Legend
par(new=TRUE, plt=c(0, 1, .9, 1))
plot(0, 0, type="n", xlim=c(0, 1), ylim=c(-.1,1), xlab="", ylab="", axes=FALSE)
rect(xleft = c(.4, .45, .5, .55, .6)-.025,
xright = c(.45, .5, .55, .6, .65)-.025,
ybottom = c(0,0,0,0,0)+.1, ytop=c(.2, .2, .2, .2, .2)+.1,
col=c("#df65b0", "#e7298a", "#ce1256", "#980043", "#67001f"),
border="#ffffff", lwd=1)
text(c(.45, .5, .55, .6)-.03, c(0,0,0,0)+.1, labels = c("10", "15", "20", "25"), pos=3, cex=.8)
```
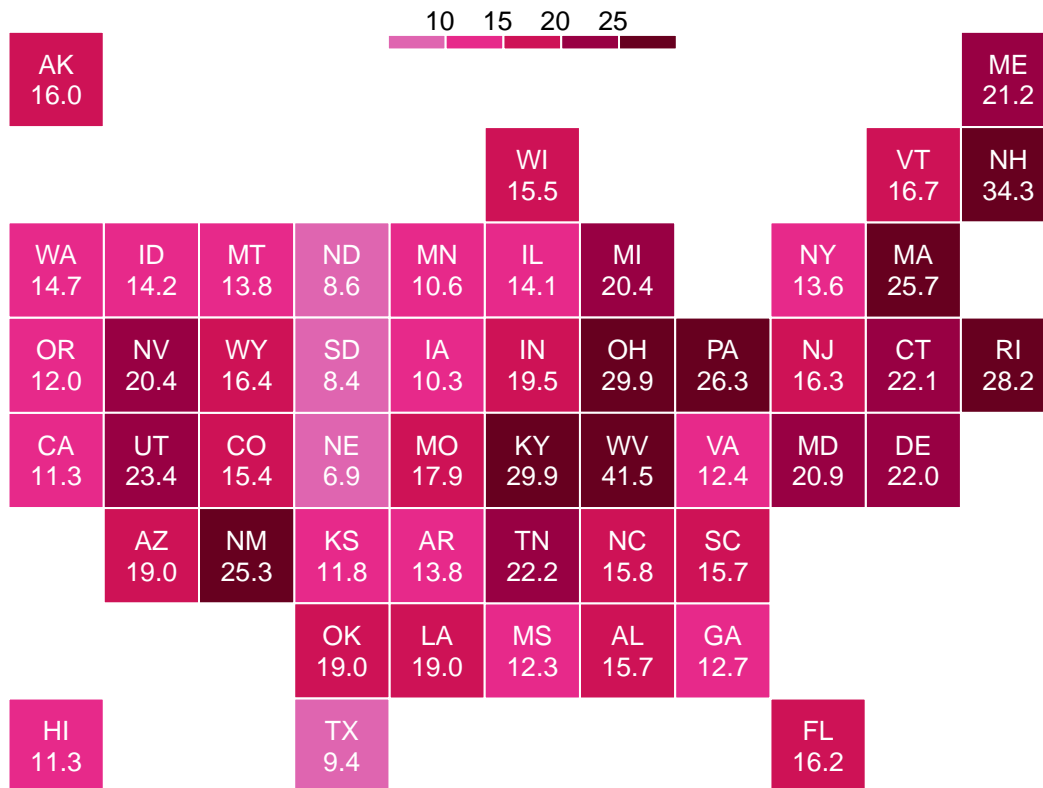
10 15 20 25

AK
14.4

ME
13.2

WI
15.0

VT
15.1

NH
15.1

WA
13.4

ID
13.4

MT
14.5

ND
2.8

MN
9.6

IL
12.1

MI
15.9

NY
11.3

MA
16.0

OR
11.3

NV
21.1

WY
17.2

SD
6.9

IA
9.3

IN
16.6

OH
20.8

PA
19.4

NJ
14.5

CT
16.0

RI
22.4

CA
11.1

UT
22.1

CO
15.5

NE
6.5

MO
17.5

KY
23.7

WV
32.2

VA
10.2

MD
14.6

DE
18.7

AZ
18.7

NM
22.6

KS
12.0

AR
11.1

TN
18.1

NC
12.9

SC
13.0

OK
20.6

LA
17.8

MS
10.8

AL
12.7

GA
10.8

HI
11.0

TX
9.3

FL
12.6

```
#Title



overdosegrid$col <- sapply(overdosegrid$Rate.2015, function(x) {
    # if (x < 5)
    #   {
    #     col <- "#d4b9da"
    # } else
      if (x < 10) {
    #     col <- "#c994c7"
    # } else if (x < 10) {
        col <- "#df65b0"
    } else if (x < 15) {
        col <- "#e7298a"
    } else if (x < 20) {
        col <- "#ce1256"
    } else if (x < 25) {
        col <- "#980043"
    } else {
        col <- "#67001f"
    }
    return(col)
})
# Start layout.
par(mar=c(0,0,0,0), bg="white")
plot(0:1, 0:1, type="n", xlab="", ylab="", axes=FALSE, asp=1)
# Draw map like before.
```

```r
par(new=TRUE, plt=c(0, 1, 0, 1))
symbols(overdosegrid$x, overdosegrid$ysideup,
        squares = rep(1, dim(overdosegrid)[1]),
        inches=FALSE,
        asp=1,
        bty="n",
        xaxt="n", yaxt="n",
        xlab="", ylab="",
        bg=overdosegrid$col,
        fg="#ffffff")
labeltext <- paste(overdosegrid$state, "\n", format(overdosegrid$Rate.2015, 2), sep="")
text(overdosegrid$x, overdosegrid$ysideup, labeltext, cex=.8, col="#ffffff")
# Legend
par(new=TRUE, plt=c(0, 1, .9, 1))
plot(0, 0, type="n", xlim=c(0, 1), ylim=c(-.1,1), xlab="", ylab="", axes=FALSE)
rect(xleft = c(.4, .45, .5, .55, .6)-.025,
xright = c(.45, .5, .55, .6, .65)-.025,
ybottom = c(0,0,0,0,0)+.1, ytop=c(.2, .2, .2, .2, .2)+.1,
col=c("#df65b0", "#e7298a", "#ce1256", "#980043", "#67001f"),
border="#ffffff", lwd=1)
text(c(.45, .5, .55, .6)-.03, c(0,0,0,0)+.1, labels = c("10", "15", "20", "25"), pos=3, cex=.8)
```



```r
#Title
```

```r
library(tidyverse)
alluv <- longset %>% dplyr::group_by(highdense, region, Significant) %>% tally()

library(alluvial)
cols <- c("#73c6b6", "#772877", "#7C821E", "#D8B98B", "#7A4012", "#c6d4e1", "#2f2016", "#fcfaea", "#456"
```

```
# cols<- c("#c6d4e1", "#2f2016", "#fcfaea", "#456789", "#772877", "#7C821E", "#D8B98B", "#2f2016","#c6d.
```

```
alluvial(alluv[1:3], freq=alluv$n, alpha=0.68, xw=0.2,cex.axis=0.8, cex = 0.6, blocks=T, border="white"
```