

heatmap

Hanna Rodrigues Ferreira

13 junho,2021

```
knitr::opts_chunk$set(echo = TRUE)
```

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(stringr)
```

```
m <- read.csv("measles_lev1.csv",header=T,stringsAsFactors=F,skip=2)
```

```
m2 <- m %>%
  # convert data to long format
  gather(key="state",
         value="value",
         -YEAR,-WEEK) %>%
  # rename columns
  setNames(c("year",
            "week",
            "state",
            "value")) %>%
  mutate(year=factor(year)) %>%
  mutate(week=factor(week)) %>%
  mutate(value=as.numeric(value))
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduzidos por coerção
```

```
# removes . and change states to title case using custom function
fn_tc <- function(x) paste(str_to_title(unlist(strsplit(x,"[.]"))),collapse=" ")
m2$state <- sapply(m2$state,fn_tc)
```

```
# custom sum function returns NA when all values in set are NA,
# in a set mixed with NAs, NAs are removed and remaining summed.
na_sum <- function(x)
{
  if(all(is.na(x))) val <- sum(x,na.rm=F)
  if(!all(is.na(x))) val <- sum(x,na.rm=T)
  return(val)
}
```

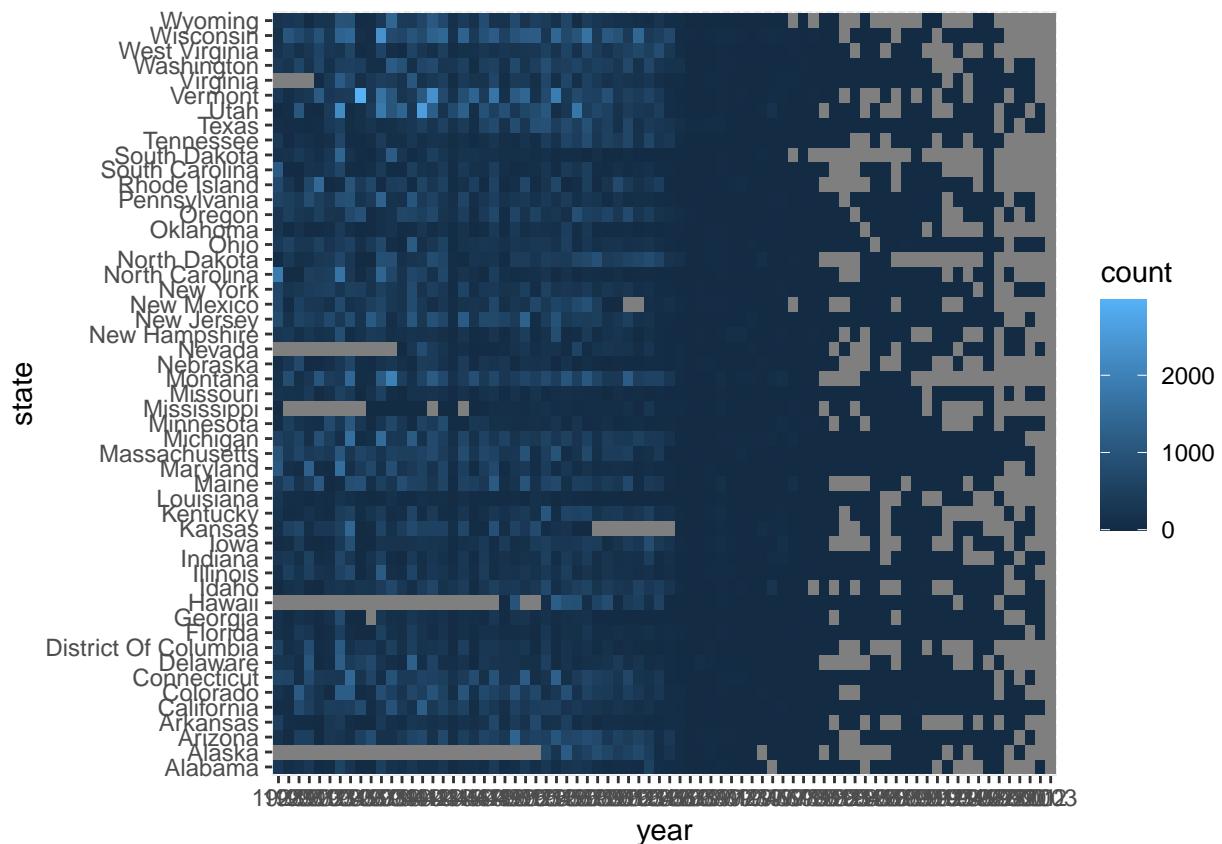
```
# sum incidences for all weeks into one year
m3 <- m2 %>%
```

```
group_by(year,state) %>%
summarise(count=na_sum(value)) %>%
as.data.frame()
```

'summarise()' has grouped output by 'year'. You can override using the '.groups' argument.

```
#basic ggplot
p <- ggplot(m3,aes(x=year,y=state,fill=count))+
  geom_tile()

p
```



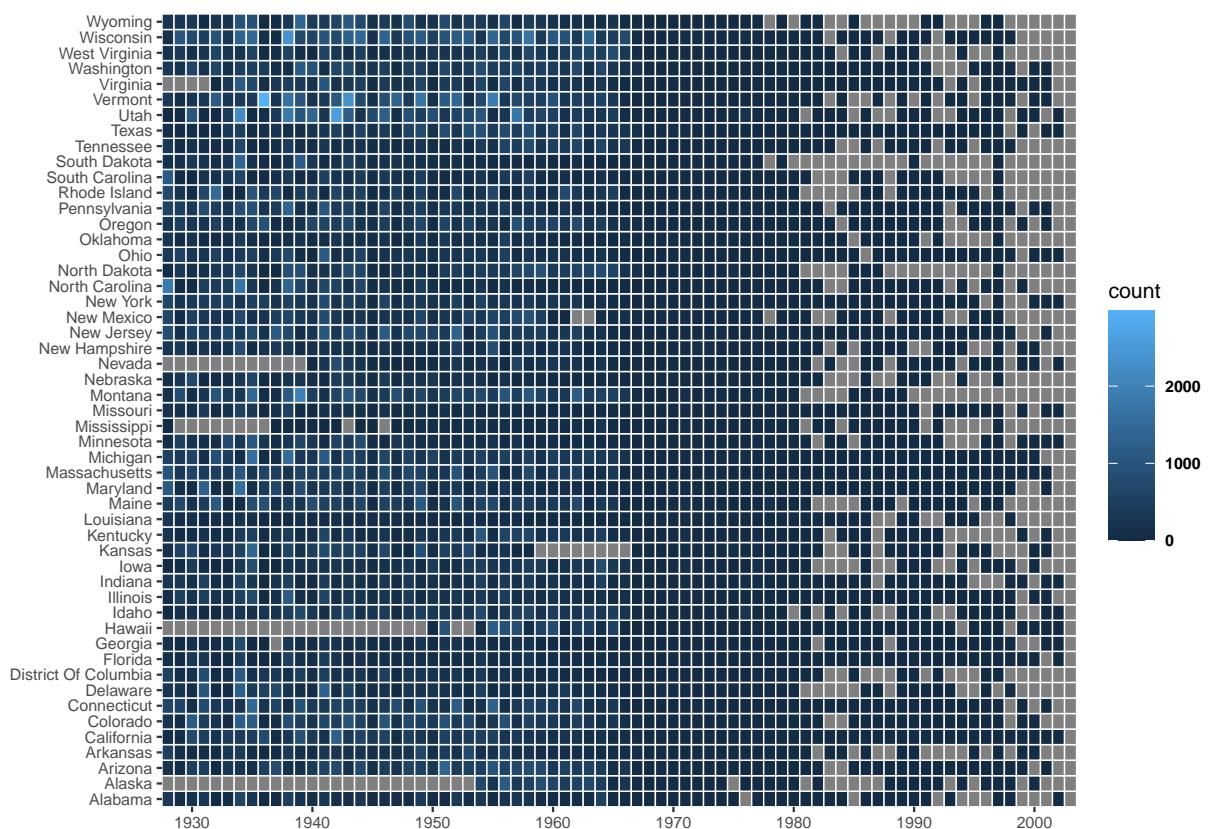
```
#modified ggplot
p <- ggplot(m3,aes(x=year,y=state,fill=count))+
  #add border white colour of line thickness 0.25
  geom_tile(colour="white",size=0.25)+
  #remove x and y axis labels
  labs(x="",y="")+
  #remove extra space
  scale_y_discrete(expand=c(0,0))+
  #define new breaks on x-axis
  scale_x_discrete(expand=c(0,0),
                    breaks=c("1930","1940","1950","1960","1970","1980","1990","2000"))+
  #set a base size for all fonts
```

```

theme_grey(base_size=8)+
#theme options
theme(
  #bold font for legend text
  legend.text=element_text(face="bold"),
  #set thickness of axis ticks
  axis.ticks=element_line(size=0.4),
  #remove plot background
  plot.background=element_blank(),
  #remove plot border
  panel.border=element_blank())

```

p



```

m4 <- m3 %>%
  # convert state to factor and reverse order of levels
  mutate(state=factor(state,levels=rev(sort(unique(state)))))) %>%
  # create a new variable from count
  mutate(countfactor=cut(count,breaks=c(-1,0,1,10,100,500,1000,max(count,na.rm=T)),
    labels=c("0","0-1","1-10","10-100","100-500","500-1000",>1000"))) %>%
  # change level order
  mutate(countfactor=factor(as.character(countfactor),levels=rev(levels(countfactor))))

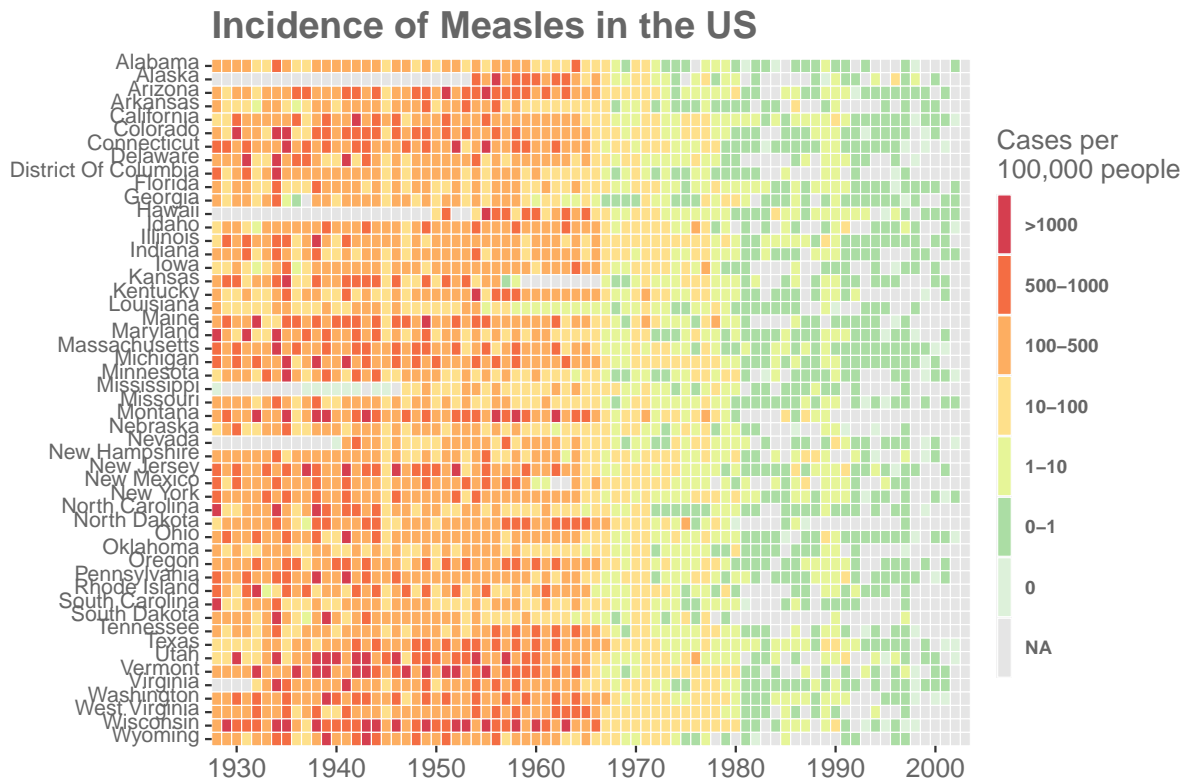
```

```

# assign text colour
textcol <- "grey40"

# further modified ggplot
p <- ggplot(m4,aes(x=year,y=state,fill=countfactor))+
  geom_tile(colour="white",size=0.2)+
  guides(fill=guide_legend(title="Cases per\n100,000 people"))+
  labs(x="",y="",title="Incidence of Measles in the US")+
  scale_y_discrete(expand=c(0,0))+
  scale_x_discrete(expand=c(0,0),breaks=c("1930","1940","1950","1960","1970","1980","1990","2000"))+
  scale_fill_manual(values=c("#d53e4f","#f46d43","#fdae61","#fee08b","#e6f598","#abdda4","#dddf1da"),na.
#coord_fixed()+
theme_grey(base_size=10)+
theme(legend.position="right",legend.direction="vertical",
      legend.title=element_text(colour=textcol),
      legend.margin=margin(grid::unit(0,"cm")),
      legend.text=element_text(colour=textcol,size=7,face="bold"),
      legend.key.height=grid::unit(0.8,"cm"),
      legend.key.width=grid::unit(0.2,"cm"),
      axis.text.x=element_text(size=10,colour=textcol),
      axis.text.y=element_text(vjust=0.2,colour=textcol),
      axis.ticks=element_line(size=0.4),
      plot.background=element_blank(),
      panel.border=element_blank(),
      plot.margin=margin(0.7,0.4,0.1,0.2,"cm"),
      plot.title=element_text(colour=textcol,hjust=0,size=14,face="bold"))
p

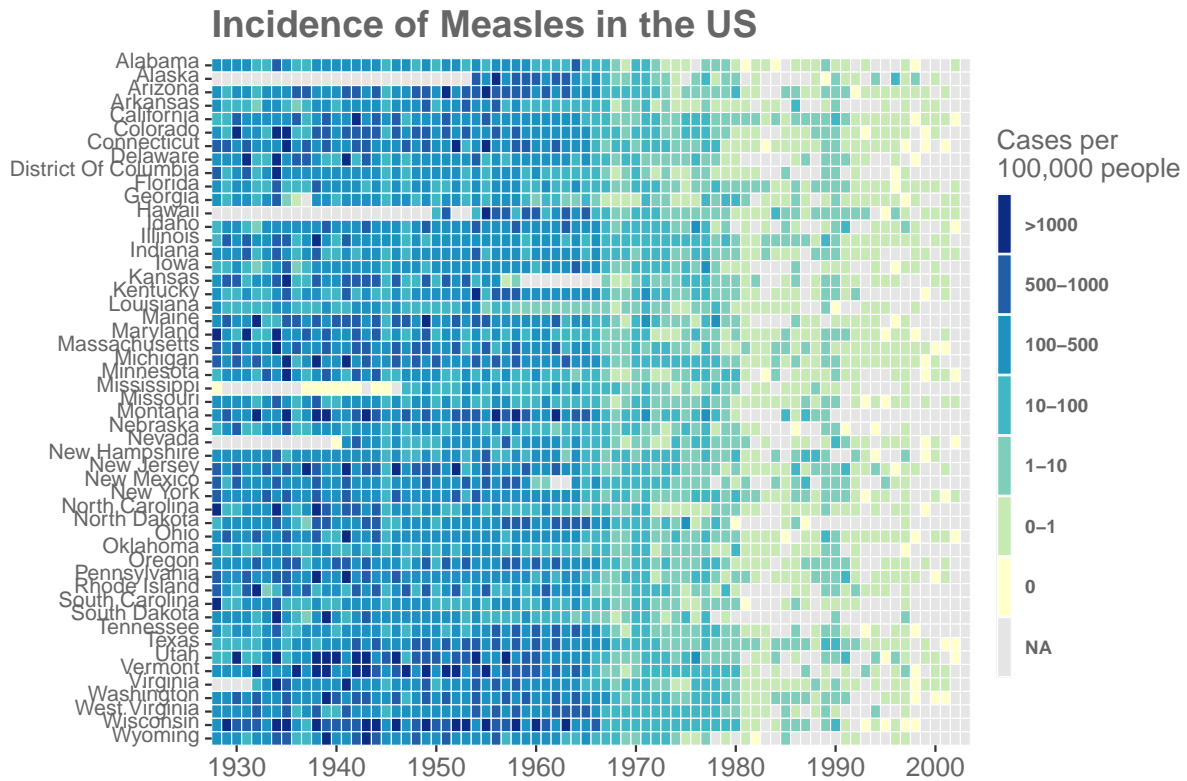
```



```
library(RColorBrewer)

# further modified ggplot
p <- ggplot(m4,aes(x=year,y=state,fill=countfactor))+
  geom_tile(colour="white",size=0.2)+
  guides(fill=guide_legend(title="Cases per\n100,000 people"))+
  labs(x="",y="",title="Incidence of Measles in the US")+
  scale_y_discrete(expand=c(0,0))+
  scale_x_discrete(expand=c(0,0),breaks=c("1930","1940","1950","1960","1970","1980","1990","2000"))+
  scale_fill_manual(values=rev(brewer.pal(7,"YlGnBu")),na.value="grey90")+
  theme_grey(base_size=10)+
  theme(legend.position="right",legend.direction="vertical",
        legend.title=element_text(colour=textcol),
        legend.margin=margin(grid::unit(0,"cm")),
        legend.text=element_text(colour=textcol,size=7,face="bold"),
        legend.key.height=grid::unit(0.8,"cm"),
        legend.key.width=grid::unit(0.2,"cm"),
        axis.text.x=element_text(size=10,colour=textcol),
        axis.text.y=element_text(vjust=0.2,colour=textcol),
        axis.ticks=element_line(size=0.4),
        plot.background=element_blank(),
        panel.border=element_blank(),
        plot.margin=margin(0.7,0.4,0.1,0.2,"cm"),
        plot.title=element_text(colour=textcol,hjust=0,size=14,face="bold"))
```

p



```
# load package
library(gplots) # heatmap.2() function
```

```
## Warning: package 'gplots' was built under R version 4.0.5
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
## lowess
```

```
library(plotrix) # gradient.rect() function
```

```
##
## Attaching package: 'plotrix'
```

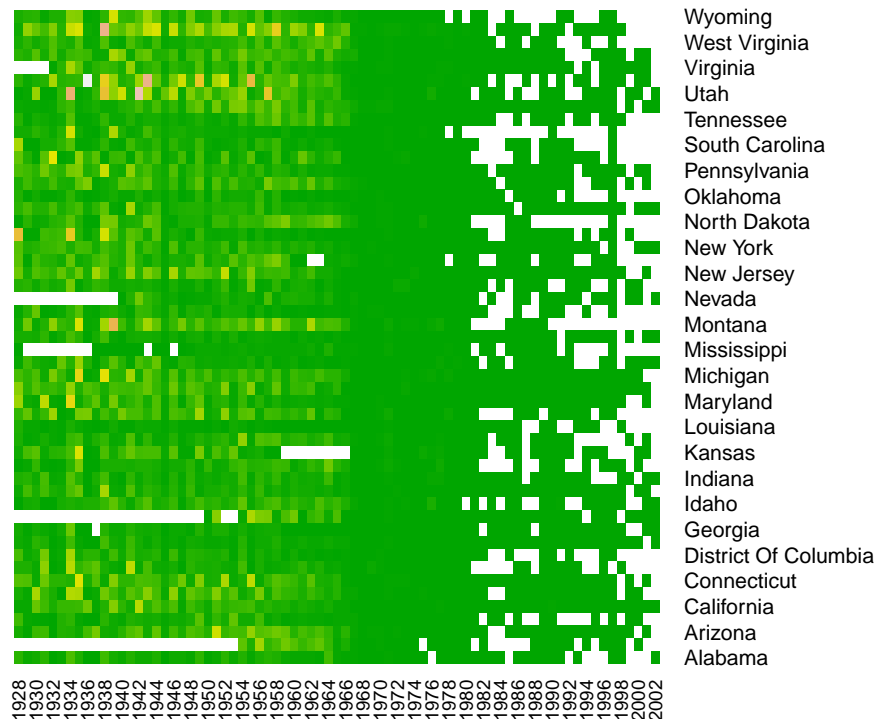
```
## The following object is masked from 'package:gplots':
##
## plotCI
```

```
# convert from long format to wide format
m5 <- m3 %>% spread(key="state",value=count)
m6 <- as.matrix(m5[,-1])
rownames(m6) <- m5$year

#base heatmap

heatmap(t(m6),Rowv=NA,Colv=NA,na.rm=T,scale="none",col=terrain.colors(100),
  xlab="",ylab="",main="Incidence of Measles in the US")
```

Incidence of Measles in the US



```
# gplots heatmap.2

par(mar=c(2,3,3,2))
gplots::heatmap.2(t(m6),na.rm=T,dendrogram="none",Rowv=NULL,Colv="Rowv",trace="none",scale="none",offset=
  breaks=c(-1,0,1,10,100,500,1000,max(m4$count,na.rm=T)),colsep=which(seq(1928,2003)%%1000),
  margin=c(3,8),col=rev(c("#d53e4f","#f46d43","#fdae61","#fee08b","#e6f598","#abdda4","#5dade2")),
  xlab="",ylab="",key=F,lhei=c(0.1,0.9),lwid=c(0.2,0.8))
gradient.rect(0.125,0.25,0.135,0.75,nslices=7,border=F,gradient="y",col=rev(c("#d53e4f","#f46d43","#fdae61",
text(x=rep(0.118,7),y=seq(0.28,0.72,by=0.07),adj=1,cex=0.8,labels=c("0","0-1","1-10","10-100","100-500","500-1000"),
text(x=0.135,y=0.82,labels="Cases per\n100,000 people",adj=1,cex=0.85)
title(main="Incidence of Measles in the US",line=1,oma=T,adj=0.21)
```

Incidence of Measles in the US

