

heatmapping in R

- b.hogan@snhu.edu
- <https://github.com/bbe2/data/blob/master/streetsweeping-citations-2018-clean.7z>

```
library(tidyr)
library(ellipse, warn.conflicts = FALSE, quietly = TRUE)
library(RColorBrewer, warn.conflicts = FALSE, quietly = TRUE)
library(gdata, warn.conflicts = FALSE, quietly = TRUE)
library(sqldf, warn.conflicts = FALSE, quietly = TRUE)
library(ggplot2, warn.conflicts = FALSE, quietly = TRUE)
library(reshape2, warn.conflicts = FALSE, quietly = TRUE)
library(ggmap, warn.conflicts = FALSE, quietly = TRUE)

#data = https://github.com/bbe2/data/blob/master/streetsweeping-citations-2018-clean.7z

df0 <- read.csv("C:\\Users\\17574\\Desktop\\data\\streetsweeping-citations-2018.csv")
x <- apply(apply(df0, 2, is.na), 2, sum) #confirm no NAs...
df1 <- data.frame(df0) ##594546 obs. of 27 variables:

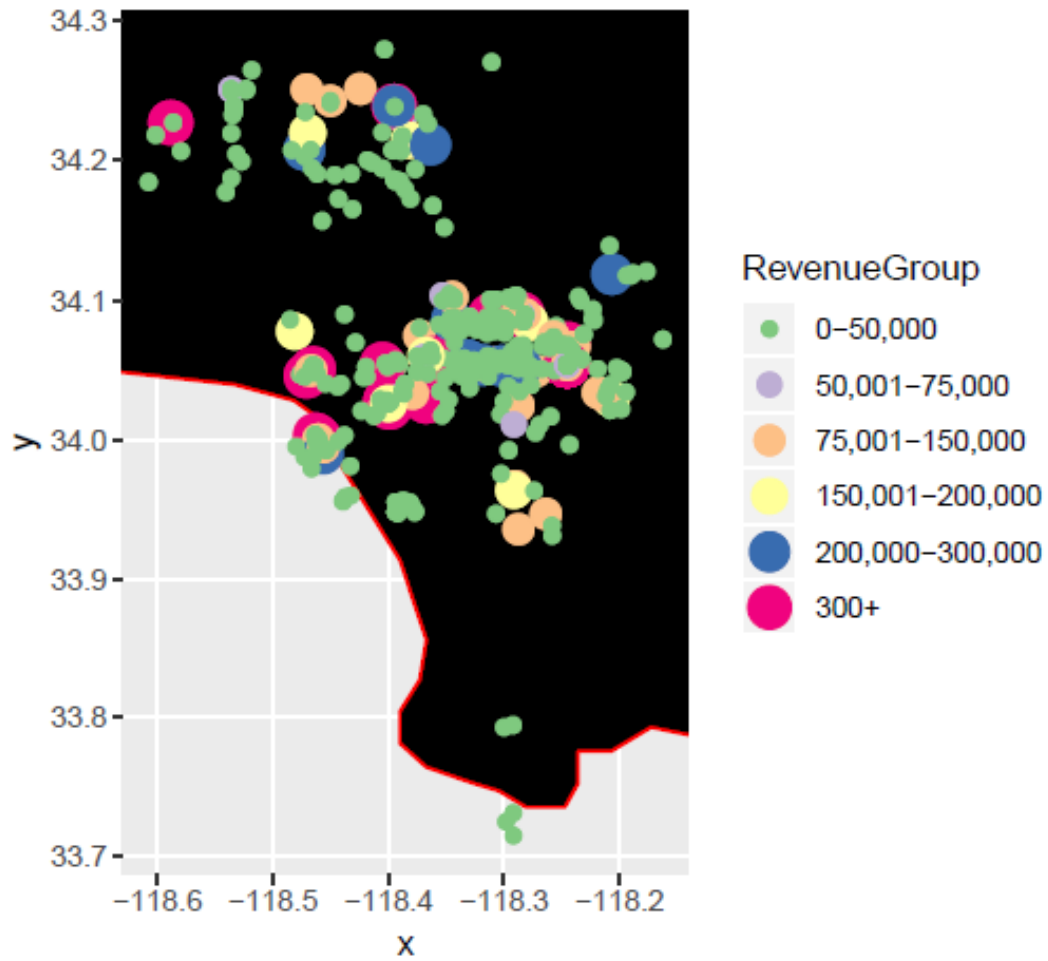
colnames(df1) <- c.names.2
remove(c.names.2)
#route.ID ==> levels=674
df.temp <- data.frame(unique(df1$routeid))
level.rows <- c(1:nrow(df.temp)) #fill in blanks
df.temp <- cbind(df.temp, level.rows) #lookup table
colnames(df.temp) <- c("routeid", paste("routeid2"))
df1 <- merge(x=df1, y=df.temp, by="routeid", all.x=FALSE)
#plate.state ==> 7levels=73
df.temp <- data.frame(unique(df1$platestate))
level.rows <- c(1:nrow(df.temp)) #fill in blanks
df.temp <- cbind(df.temp, level.rows) #lookup table
colnames(df.temp) <- c("platestate", paste("platestate2"))
df1 <- merge(x=df1, y=df.temp, by="platestate", all.x=FALSE)
#car.make ==> levels=62
df.temp <- data.frame(unique(df1$carmake))
level.rows <- c(1:nrow(df.temp)) #fill in blanks
df.temp <- cbind(df.temp, level.rows) #lookup table
colnames(df.temp) <- c("carmake", paste("carmake2"))
df.carid.master <- df.temp
df1 <- merge(x=df1, y=df.temp, by="carmake", all.x=FALSE)
write.csv(df.carid.master, "df.carid.master.csv") #sped up normalization
#car.bodystyle ==> levels=12
df.temp <- data.frame(unique(df1$carbodystyle))
level.rows <- c(1:nrow(df.temp)) #fill in blanks
df.temp <- cbind(df.temp, level.rows) #lookup table
colnames(df.temp) <- c("carbodystyle", paste("carbodystyle2"))
df1 <- merge(x=df1, y=df.temp, by="carbodystyle", all.x=TRUE)
#car.color ==> levels=16
df.temp <- data.frame(unique(df1$carcolor))
level.rows <- c(1:nrow(df.temp)) #fill in blanks
df.temp <- cbind(df.temp, level.rows) #lookup table
colnames(df.temp) <- c("carcolor", paste("carcolor2"))
df1 <- merge(x=df1, y=df.temp, by="carcolor", all.x=TRUE)
#issue.address ==> levels=292164
df.temp <- data.frame(unique(df1$issueaddress))
level.rows <- c(1:nrow(df.temp)) #fill in blanks
df.temp <- cbind(df.temp, level.rows) #lookup table
colnames(df.temp) <- c("issueaddress", paste("issueaddress2"))
df1 <- merge(x=df1, y=df.temp, by="issueaddress", all.x=TRUE)
#head(df1, 3)
```

```

#-----LA REVENUE MAP-----
state <- map_data("state")
route.df1 <- sqldf('select routeid2, issueaddresslat as long, issueaddresslon as lat,
SUM(violationfineamt) as fine from df1 group by routeid2')
RevenueGroup <- c(1:nrow(route.df1)) #fill in blanks to add a group size variable
state <- c(1:nrow(route.df1)) #write.csv(route.df1, "Route_Fine.csv")
longnew <- c(1:nrow(route.df1))
latnew <- c(1:nrow(route.df1))
route.df1 <- cbind(route.df1, RevenueGroup) #Gadd state full name back for mapping please
route.df1 <- cbind(route.df1, state)
route.df1 <- cbind(route.df1, longnew)
route.df1 <- cbind(route.df1, latnew)
remove(RevenueGroup, state, latnew, longnew) #----ADD FIXING OF LAT/LON CONVERSION HERE
#str(route.df1)
# there are both bad and blank coordinates; I am just grouping for revenue diagram
#old values = -137.9131, 27.51751
#new values = -118.24532, 34.05349
n <- 1
while (n <= nrow(route.df1))
{ route.df1[n,7] = round(route.df1[n,2],4)
route.df1[n,8] = round(route.df1[n,3],5)
#route.df1[n,2] = 0
#route.df1[n,3] = 0
n <- n+1 #fixing some bad lat long positions - moving over from ocean
} #REVENUE DOT PLOT FIXING LAT/LONG
n <- 1
while (n <= nrow(route.df1))
{ #if (route.df1[n,7] == -137.9131) { baddates <- baddates+1 }
if (route.df1[n,7] == -137.9131) {route.df1[n,2] = -118.24532 }
if (route.df1[n,8] == 27.51751) {route.df1[n,3] = 34.05349 }
if (route.df1[n,7] == 0) {route.df1[n,2] = -118.24532 }
if (route.df1[n,8] == 0) {route.df1[n,3] = 34.05349 }
n <- n+1 }
#colnames(route.df1) <- c("routeid2", "long", "lat", "fine", "grpsize", "state", "fix")
#head(route.df1)
route.df1$state <- "california" #expand to get whole view on R windows
route.df1$RevenueGroup <- cut(route.df1$fine, #make buckets for mapping
breaks = c(-Inf, 50000, 75000, 150000, 200000, 300000, Inf),
labels = c("0-50,000", "50,001-75,000", "75,001-150,000", "150,001-200,000", "200,000-300,000", "300+" ),
right = FALSE)
mycolors <- brewer.pal(6, "Accent") #head(df4,1)
names(mycolors) <- levels(route.df1$grpsize) #getting the color names
ggplot(route.df1, aes(map_id=state)) +
expand_limits(x=route.df1$long, y=route.df1$lat) + coord_map() +
geom_map(map = us, fill="black", color="red" ) + geom_point(data=route.df1,
aes(x=long, y=lat, color=RevenueGroup, size=RevenueGroup )) +
#stat_density2d(data=route.df1, aes(x=long, y=lat), geom="density_2d") +
scale_colour_manual( values=mycolors) + #name="Color",
ggtitle("Los Angeles Sweep Fine Revenue by 674 Routes")

```

Los Angeles Sweep Fine Revenue by 674 Routes



#----->>>> HEATMAP WORK <<<=====

```
car.df1 <- data.frame(sqldf('select carmake,SUM(carmake2) as CarQty
from df1 group by carmake'))
```

```
cargrouppname <- c("Dom.EconY","Dom.EconN","Intl.EconY","Intl.EconN","Fancy","Trucks")
```

```
cargroup<- c(1,2,3,4,5,6)
```

```
df.car.group <- data.frame(cargroup, cargrouppname)
```

```
df.car.group
```

```
## cargroup cargrouppname
```

```
## 1 1 Dom.EconY
```

```
## 2 2 Dom.EconN
```

```
## 3 3 Intl.EconY
```

```
## 4 4 Intl.EconN
```

```
## 5 5 Fancy
```

```
## 6 6 Trucks
```

```
df2.heat <-merge(x=df2.heat, y=df.car.group, by="cargroup", all.x = TRUE)
```

```
colnames(df2.heat)
```

```
## [1] "cargroup"      "carmake2"
```

```
"carmake"
```

```
## [4] "issueaddress"  "carcolor"
```

```
"carbodystyle"
```

```
## [7] "platestate"    "routeid"
```

```
"X"
```

```
## [10] "ticketnumber"  "issuedate"
```

```
"issueyear"
```

```
## [13] "issuemonth"    "issueday"
```

```
"issueweekday"
```

```
## [16] "issuetime"     "issuetimebin"
```

```
"agencyid"
```

```
## [19] "meterid"       "issueaddresslat"
```

```
"issueaddresslon"
```

```
## [22] "violationid"   "violationdesc"
```

```
"violationfineamt"
```

```
## [25] "plateexpiredate" "plateexpireyear"
```

```
"plateexpiremonth"
```

```
## [28] "plateexpireflag" "carmakeimportflag"
```

```
"routeid2"
```

```
## [31] "platestate2"    "carbodystyle2"
```

```
"carcolor2"
```

```
## [34] "issueaddress2"  "carqty"
```

```
"cargrouppnormalize"
```

```
## [37] "cargrouppname"
```

#Step 4 Heatmap---using normalization approach between 0-1 as didnt' want negative scale

```
df3.heat <- data.frame(df2.heat[,c(9,12,13,15,17,1,37,36)])
```

```
#head(df3.heat,1)
```

```
max(df3.heat$issuetimebin)
```

```
## [1] 6
```

```
range(df3.heat$issuetimebin)
```

```
## [1] 0 6
```

```
df3.heat$timebin.z <- (df3.heat[,5]-min(df3.heat$issuetimebin))/
```

```
(max(df3.heat$issuetimebin)-min(df3.heat$issuetimebin))
```

```
df3.heat$cargrp.z <- (df3.heat[,6]-min(df3.heat$cargroup))/
```

```
(max(df3.heat$cargroup)-min(df3.heat$cargroup))
```

```
# head(df3.heat,1)
```

```
# length(df3.heat)
```

```
dftemp <-data.frame(df3.heat) # <----switch to dataframe w no dots in it
```

```
#-----
```

#hwat I really want is by car groups.....

```
cargrp <- sqldf('select issuemonth as month, issueweekday as day,
```

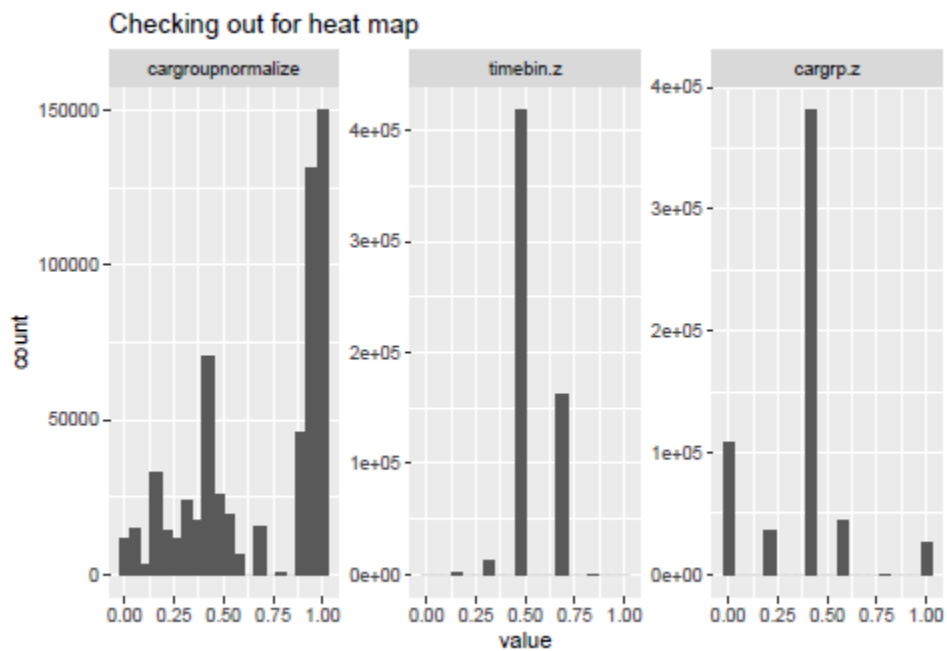
```
cargrouppname , cargrouppnormalize as zscore from dftemp')
```

#Comparison to learn if need to transform data to interpret better in heatmap

```
ggplot(data = melt(df3.heat[,c(8:10)]), mapping = aes(x = value)) + geom_histogram(bins=20)+
```

```
facet_wrap(~variable, scales = "free") +
```

```
ggtitle("Checking out for heat map") ## No id variables; using all as measure
```



#Abundance = grouping of all air quality factor values 0-1
#Felt Square Root Transformation did help with graph read but could vary by person
 df4.heat <- **gather**(data = cargrp, key = Class, value= Abundance,-c(1:3))
 df4.heat\$Sqrt.Abandance <- **sqr**t(df4.heat\$Abundance)
 heat.reg <- **ggplot**(data=df4.heat, mapping= **aes**(x=day,y=cargroupname,fill=Abundance))+
geom_tile() + **xlab**(label="Month/Day") + **ylab**(label="ticket") +
ggtitle("Super Luxury Cars e.g. Rolls Royces & Bentleys towed Least of Others but Still 3% or 21,600 of
theme(axis.text.x = **element_text**(size=8,angle = 90, hjust = 1)) +
facet_grid(~ month, switch = "x", scales="free_x", space="free_x")
 heat.reg *#non transformed data*

