HW2 Creative R Plot for SWARM

NA

```
library(foreign)
library(ggplot2)
library(magrittr)
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.4.4
```

```
# use file.choose() to locate the file
NA

# have a general idea of the dataset
str(nyc_locpro)
```

```
## 'data.frame':    579 obs. of  6 variables:
##  $ user_id       : int  7888379 139361183 30911792 68325822 63382043 33239212 5329
8833 767623 2815399 6289088 ...
##  $ if_local      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ gender        : Factor w/ 3 levels "female","male",..: 2 2 3 2 1 2 2 2 1 3 ...
##  $ checkins_count: int  1683 313 448 1066 1772 470 4260 7098 794 1448 ...
##  $ photos_count  : int  25 3 52 46 159 0 150 84 221 160 ...
##  $ friends_count : int  89 2 92 120 4 278 25 122 110 138 ...
```

```
summary(nyc_locpro)
```

```
##     user_id             if_local    gender    checkins_count
##  Min.   :     4556   Min.   :1   female:163   Min.   :    3
##  1st Qu.:  4540591   1st Qu.:1   male  :403   1st Qu.: 1324
##  Median : 19353883   Median :1   none  : 13   Median : 2530
##  Mean   : 25294696   Mean   :1                Mean   : 4110
##  3rd Qu.: 39918697   3rd Qu.:1                3rd Qu.: 4962
##  Max.   :153634738   Max.   :1                Max.   :51173
##   photos_count    friends_count
##  Min.   :   0.0   Min.   :  1.0
##  1st Qu.:  11.0   1st Qu.: 42.5
##  Median :  38.0   Median : 85.0
##  Mean   : 166.6   Mean   :161.6
##  3rd Qu.: 127.5   3rd Qu.:204.5
##  Max.   :4684.0   Max.   :997.0
```

```
Op_nyc_locpro <- subset(nyc_locpro,select = c("checkins_count","photos_count","friend
s_count"))
str(Op_nyc_locpro)
```

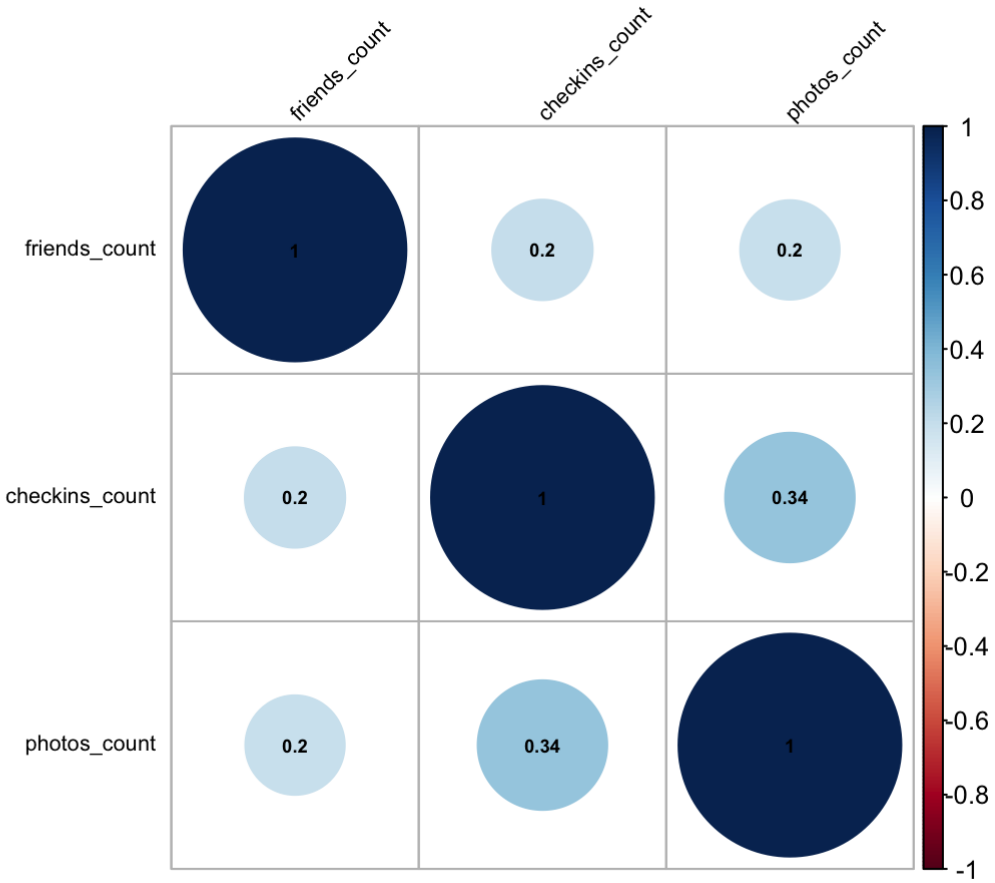
```
## 'data.frame':    579 obs. of  3 variables:
##  $ checkins_count: int  1683 313 448 1066 1772 470 4260 7098 794 1448 ...
##  $ photos_count  : int  25 3 52 46 159 0 150 84 221 160 ...
##  $ friends_count : int  89 2 92 120 4 278 25 122 110 138 ...
```

```
# correlation matrix for 3 variables
Op_corr <- cor(Op_nyc_locpro, method = "pearson", use = "complete.obs")
NA

library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.4.2
```

```
## corrplot 0.84 loaded
```

```
corrplot(Op_corr, type = "full", order = "hclust", tl.col = "black", tl.srt = 45,tl.cex = .68,addCoef.col="black", number.cex = .58)
```




```
# with distribution
library(PerformanceAnalytics)
```

```
## Warning: package 'PerformanceAnalytics' was built under R version 3.4.3
```

```
## Loading required package: xts
```

```
## Warning: package 'xts' was built under R version 3.4.4
```

```
## Loading required package: zoo
```

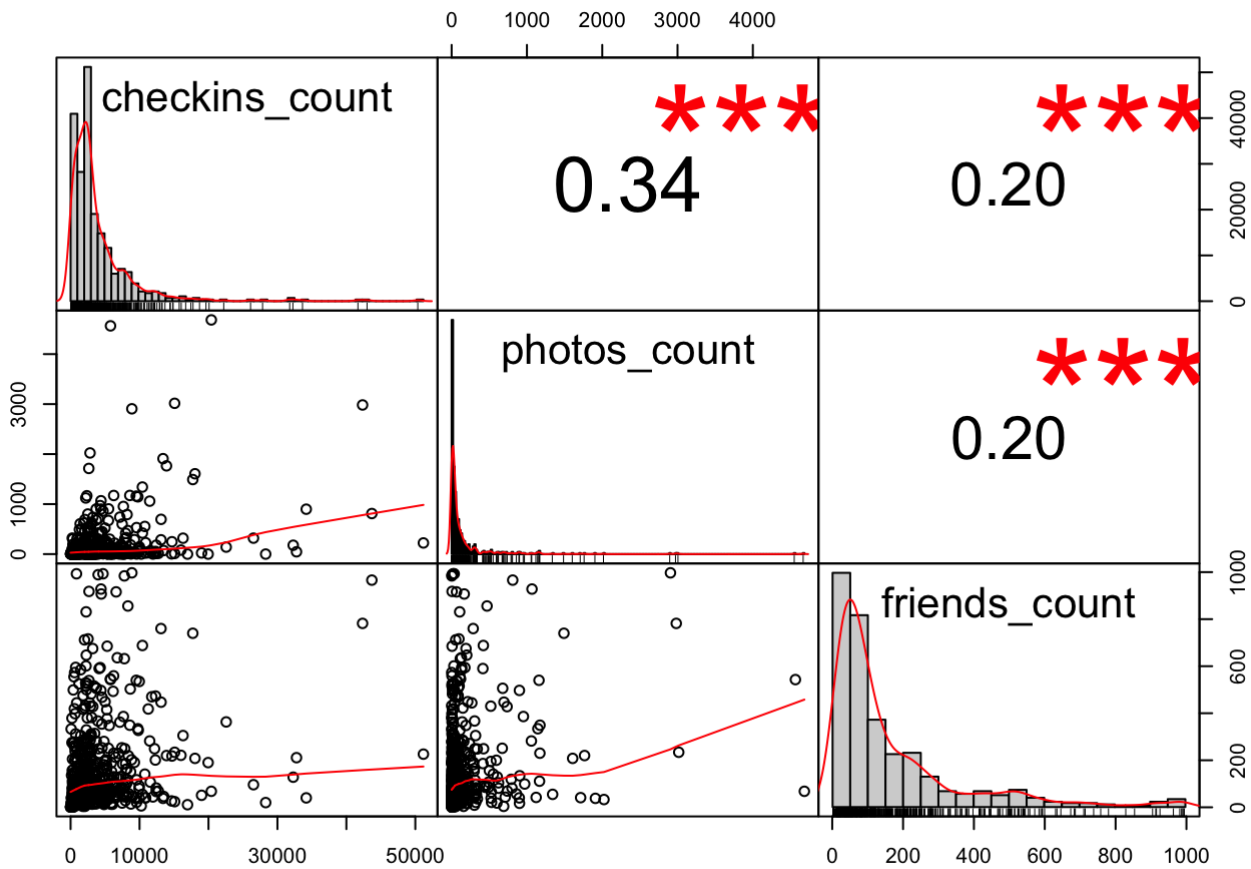
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
##
##     legend
```

```
chart.Correlation(Op_nyc_locpro, histogram=TRUE, pch=19, text.col = 'b')
```



NA

```
# use file.choose() to locate the file
NA
NA
str(nyc_locchk)
```

```
## 'data.frame':    593557 obs. of  8 variables:
##  $ user_id      : int  10008414 10008414 10008414 10008414 10008414 10008414 10008
414 10008414 10008414 10008414 ...
##  $ if_local     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ venue_id     : Factor w/ 48397 levels "3fd66200f964a52000e71ee3",..: 1497 23103
1124 18070 36890 27541 1124 8371 20909 1687 ...
##  $ category_name: Factor w/ 629 levels "Accessories Stores",..: 302 94 427 205 46
80 427 372 483 246 ...
##  $ category_id  : Factor w/ 628 levels "4bf58dd8d48988d100941735",..: 302 150 140
67 44 459 140 308 355 25 ...
##  $ lat          : num  40.7 40.7 40.8 40.8 40.7 ...
##  $ lng          : num  -74 -74 -74 -74 -74 ...
##  $ created_time : Factor w/ 440245 levels "2012-12-31 19:03",..: 320668 320199 320
163 320067 319642 319625 319502 318688 318670 318657 ...
```

```
time <- nyc_locchk$created_time
# time <- as.character(time)
NA

# interactive R plot
# Library
library(dygraphs)
```

```
## Warning: package 'dygraphs' was built under R version 3.4.4
```

```
library(xts)          # To make the convertion data-frame / xts format
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.2
```

```
NA
```

```
NA
NA
NA
NA
```

```
## Warning: package 'tibble' was built under R version 3.4.3
```

```
## Warning: package 'tidyr' was built under R version 3.4.4
```

```
## Warning: package 'purrr' was built under R version 3.4.4
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
## Warning: package 'forcats' was built under R version 3.4.3
```

```
NA
NA
NA
NA
NA
NA
NA
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.4.4
```

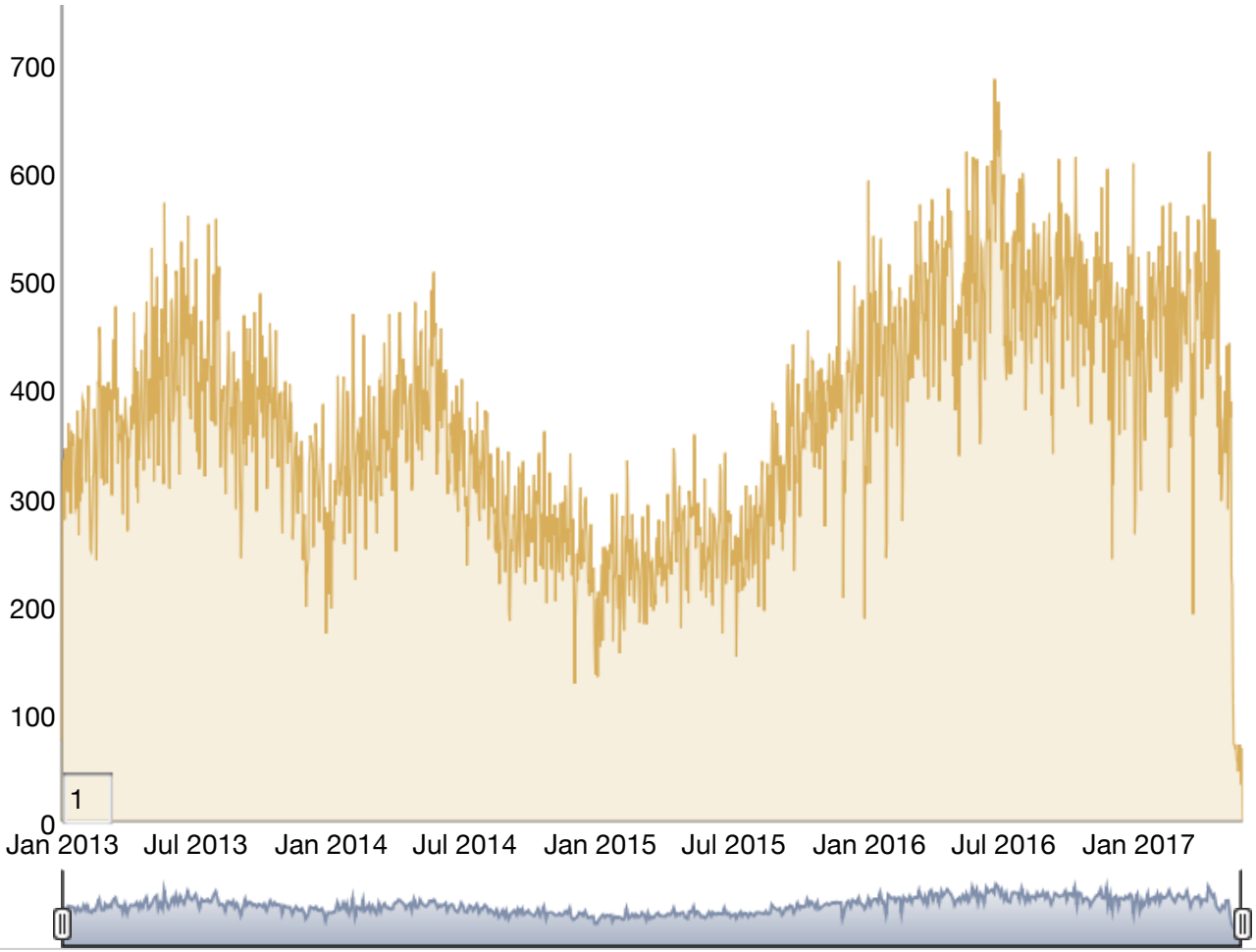
```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

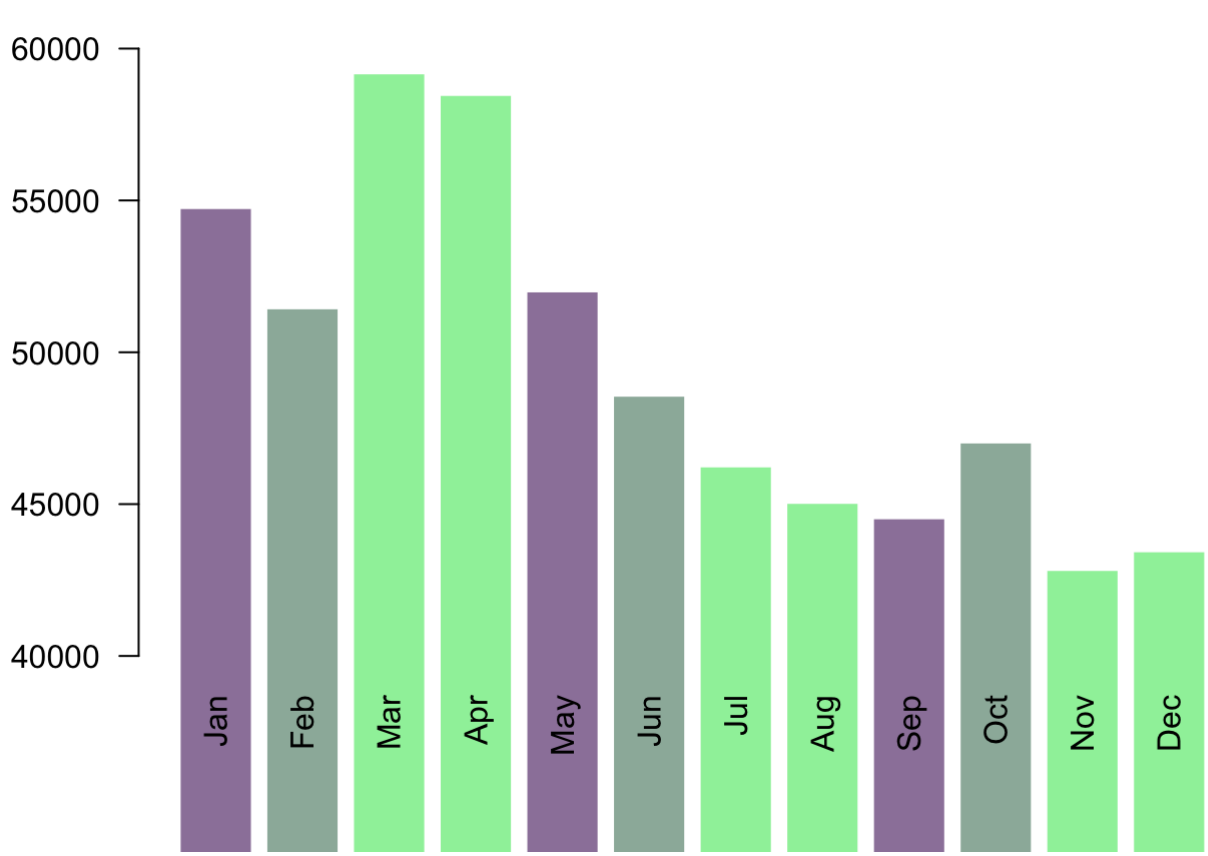
```
NA
NA
str(date_count)
```

```
## 'data.frame':    1602 obs. of  2 variables:
##  $ date: Factor w/ 1602 levels "2012-12-31","2013-01-01",..: 1 2 3 4 5 6 7 8 9 10
...
##  $ Freq: int  75 226 323 333 335 279 345 320 323 305 ...
```

```
date_fac <- date_count$date
# convert it to a date-time format
# date_fac <- ymd_hms(date_fac)
date_count$date <- as.Date(date_fac)
don=xts(x = date_count$Freq , order.by = date_count$date)
dygraph(don) %>%
  dyOptions(labelsUTC = TRUE,fillAlpha=.2, fillGraph=TRUE, drawGrid = FALSE,colors="#
D8AE5A") %>%
  # dygraph interactive range selection and zooming
  dyRangeSelector() %>%
  # draws a crosshair line over the point closest to the mouse
  dyCrosshair(direction = "both") %>% # vertical + horizontal
  # mouse-over highlighting
  dyHighlight(highlightCircleSize = 5, highlightSeriesBackgroundAlpha = 0.2, hideOnMo
useOut = FALSE)  %>%
  # dygraph rolling average period text box
  dyRoller(rollPeriod = 1)
```




```
# Month Count & Weeks daily count
month <- substr(time[],6,7)
month_count <- as.data.frame(table(month))
month_name <- c("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","De
c")
my_bar <- barplot(month_count$Freq , border=F , names.arg=month_name, las=2 , col=c(r
gb(0.3,0.1,0.4,0.6) , rgb(0.3,0.5,0.4,0.6) , rgb(0.3,0.9,0.4,0.6) ,  rgb(0.3,0.9,0.4,
0.6)) , ylim = c(40000,60000), main="NewYorkers' Check-in Plot" )
```



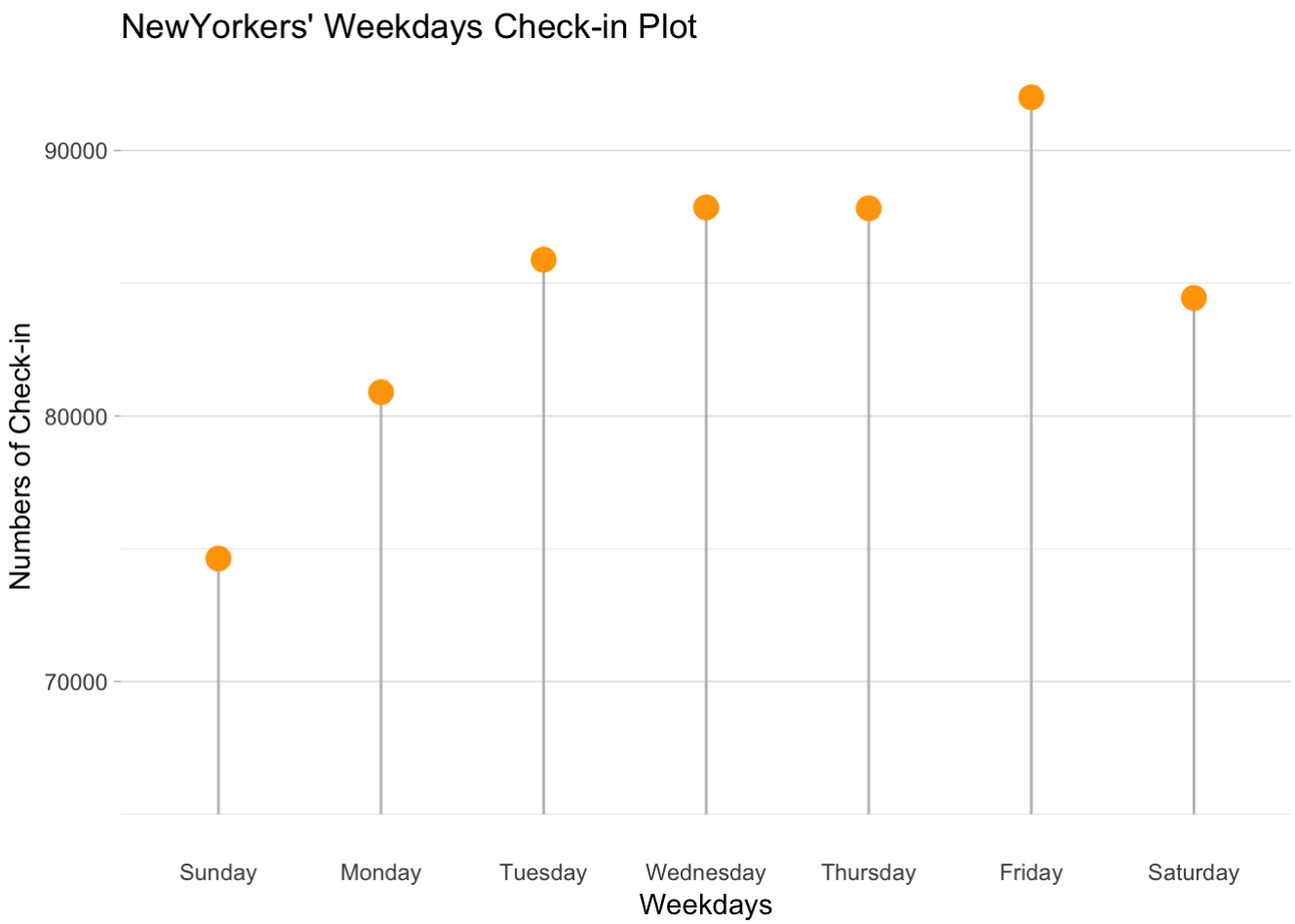

```
library(ggplot2)
date <- substr(time[],1,10)
date <- as.Date(date)
weekday <- weekdays(date)
weekday_count <- as.data.frame(table(weekday))
order(weekday_count[,2])
```

```
## [1] 4 2 3 6 5 7 1
```

```
# sort by weekdays
weekday_count$weekday <- factor(weekday_count$weekday, levels= c("Sunday", "Monday",
    "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday"))
weekday_count[order(weekday_count$weekday), ]
```

```
##     weekday  Freq
## 4    Sunday 74632
## 2    Monday 80898
## 6   Tuesday 85890
## 7 Wednesday 87857
## 5  Thursday 87826
## 1    Friday 92007
## 3  Saturday 84447
```

```
# plot
ggplot(weekday_count,aes(x=weekday, y=Freq)) +
  geom_segment(aes(x=weekday, xend=weekday, y=65000, yend=Freq),color="grey") +
  geom_point( color="orange", size=4) +
  theme_light() +
  theme(
    panel.grid.major.x = element_blank(),
    panel.border = element_blank(),
    axis.ticks.x = element_blank()
  ) +
  xlab("Weekdays") +
  ylab("Numbers of Check-in") +
  ggtitle("NewYorkers' Weekdays Check-in Plot")
```



NA

```
# Plot the map
# 1 Download the shapefile.
# 2 Read this shape file with the rgdal library.
library(rgdal)
```

```
## Warning: package 'rgdal' was built under R version 3.4.4
```

```
## Loading required package: sp
```

```
## rgdal: version: 1.3-6, (SVN revision 773)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.1.3, released 2017/20/01
##  Path to GDAL shared files: /Library/Frameworks/R.framework/Versions/3.4/Resource
s/library/rgdal/gdal
##  GDAL binary built with GEOS: FALSE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: /Library/Frameworks/R.framework/Versions/3.4/Resourc
es/library/rgdal/proj
##  Linking to sp version: 1.3-1
```

```
library(yarrr) # for creating transparent scatterplot
```

```
## Loading required package: jpeg
```

```
## Loading required package: BayesFactor
```

```
## Warning: package 'BayesFactor' was built under R version 3.4.4
```

```
## Loading required package: coda
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:tidyr':
##
##     expand
```

```
## ************
## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard M
orey (richarddmorey@gmail.com).
##
## Type BFManual() to open the manual.
## ************
```

```
## Loading required package: circlize
```

```
## Warning: package 'circlize' was built under R version 3.4.4
```

```
## ========================================
## circlize version 0.4.4
## CRAN page: https://cran.r-project.org/package=circlize
## Github page: https://github.com/jokergoo/circlize
## Documentation: http://jokergoo.github.io/circlize_book/book/
##
## If you use it in published research, please cite:
## Gu, Z. circlize implements and enhances circular visualization
##   in R. Bioinformatics 2014.
## ========================================
```

```
## yarrr v0.1.5. Citation info at citation('yarrr'). Package guide at yarrr.guide()
```

```
## Email me at Nathaniel.D.Phillips.is@gmail.com
```

```
##
## Attaching package: 'yarrr'
```

```
## The following object is masked from 'package:ggplot2':
##
##     diamonds
```

```
NA
```

```
## OGR data source with driver: ESRI Shapefile
NA
## with 5 features
## It has 4 fields
```

```
# -- > Now you have a Spdf object (spatial polygon data frame). You can start doing m
aps!
# Basic plot of this shape file:
par(mar=c(0,0,0,0))
# https://color.adobe.com/zh/explore/?filter=most-popular&time=month using this cite
 to find your color
plot(my_spdf, col="#87AEB4", fill=TRUE, bg = "#CDD4CA", lwd=0.35, mar=rep(0,4), borde
r=1 )
```

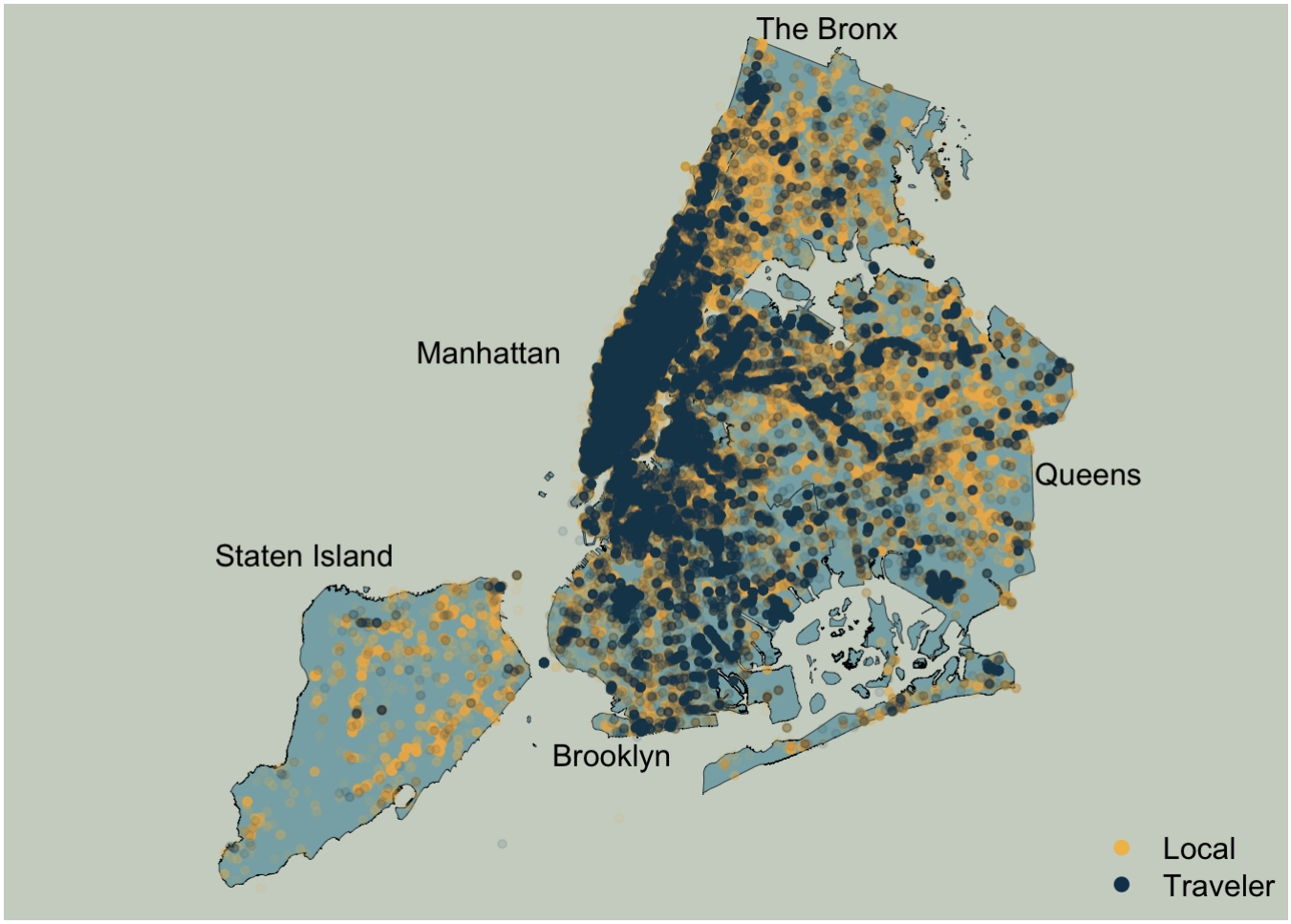
```
## Warning in title(...): "fill" is not a graphical parameter
```

```
## Warning in polypath(x = mcrds[, 1], y = mcrds[, 2], border = border, col =
## col, : "fill" is not a graphical parameter

## Warning in polypath(x = mcrds[, 1], y = mcrds[, 2], border = border, col =
## col, : "fill" is not a graphical parameter

## Warning in polypath(x = mcrds[, 1], y = mcrds[, 2], border = border, col =
## col, : "fill" is not a graphical parameter

## Warning in polypath(x = mcrds[, 1], y = mcrds[, 2], border = border, col =
## col, : "fill" is not a graphical parameter

## Warning in polypath(x = mcrds[, 1], y = mcrds[, 2], border = border, col =
## col, : "fill" is not a graphical parameter
```

```
str(nyc_locchk)
```

```
## 'data.frame':    593557 obs. of  8 variables:
##  $ user_id      : int  10008414 10008414 10008414 10008414 10008414 10008414 10008
414 10008414 10008414 10008414 ...
##  $ if_local     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ venue_id     : Factor w/ 48397 levels "3fd66200f964a52000e71ee3",..: 1497 23103
1124 18070 36890 27541 1124 8371 20909 1687 ...
##  $ category_name: Factor w/ 629 levels "Accessories Stores",..: 302 94 427 205 46
80 427 372 483 246 ...
##  $ category_id  : Factor w/ 628 levels "4bf58dd8d48988d100941735",..: 302 150 140
67 44 459 140 308 355 25 ...
##  $ lat          : num  40.7 40.7 40.8 40.8 40.7 ...
##  $ lng          : num  -74 -74 -74 -74 -74 ...
##  $ created_time : Factor w/ 440245 levels "2012-12-31 19:03",..: 320668 320199 320
163 320067 319642 319625 319502 318688 318670 318657 ...
```

```
points(nyc_locchk$lng, nyc_locchk$lat, col = transparent("#F2BE54",0.9), cex = .6, pc
h = 19)
points(nyc_trachk$lng, nyc_trachk$lat, col = transparent("#153E5C",0.9), cex = .6, pc
h = 19)
text(-74.2,40.66,'Staten Island',cex = 1)
text(-74.00,40.56,'Brooklyn',cex = 1)
text(-73.86,40.92,'The Bronx',cex = 1)
text(-74.08,40.76,'Manhattan',cex = 1)
text(-73.69,40.7,'Queens',cex = 1)
# title(line = -16,'Newyorker & Traveler Check-in Plot',cex = 1.5,outer = TRUE)
legend("bottomright",legend = c("Local","Traveler"),pch = 19,cex = 1, col = c("#F2BE5
4","#153E5C"),bty = "n")
```


The Bronx
Manhattan
Queens
Staten Island
Brooklyn
Local
Traveler