

Final Project

Ellie Carpenter, Sarina Barot-Martinez, Mihir Harshe, Rubeno Dechua, Hans Clough

2022-12-06

```
install.packages("tidyverse", repos = 'http://cran.us.r-project.org')
```

```
##  
## The downloaded binary packages are in  
## /var/folders/0f/mlslmtm962144xtv_vx74m100000gn/T//RtmplhPoiu/downloaded_packages
```

```
library(tidyverse)
```

```
## — Attaching packages — tidyverse 1.3.2 —  
## ✓ ggplot2 3.4.0      ✓ purrr 0.3.4  
## ✓ tibble 3.1.8      ✓ dplyr 1.0.10  
## ✓ tidyr 1.2.1       ✓ stringr 1.4.1  
## ✓ readr 2.1.2       ✓ forcats 0.5.2  
## — Conflicts — tidyverse_conflicts() —  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()
```

```
install.packages("ggplot", repos = 'http://cran.us.r-project.org')
```

```
## Warning: package 'ggplot' is not available for this version of R  
##  
## A version of this package for your version of R might be available elsewhere,  
## see the ideas at  
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
library(ggplot2)  
install.packages("ggpubr", repos='http://cran.us.r-project.org')
```

```
##  
## The downloaded binary packages are in  
## /var/folders/0f/mlslmtm962144xtv_vx74m100000gn/T//RtmplhPoiu/downloaded_packages
```

```
library(ggpubr)
```

```
install.packages("Rtools", repos='http://cran.us.r-project.org')
```

```
## Warning: package 'Rtools' is not available for this version of R
##
## A version of this package for your version of R might be available elsewhere,
## see the ideas at
## https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

```
install.packages("ggplot2", repos = "http://cran.rstudio.org")
```

```
##
## The downloaded binary packages are in
## /var/folders/0f/mlslmtm962144xtv_vx74m100000gn/T//RtmplhPoiu/downloaded_packages
```

```
library(ggplot2)
```

```
SFtransit <- read.csv("/Users/ellie/Downloads/Muni_Simple_Routes.csv")
str(SFtransit)
```

```
## 'data.frame':    139 obs. of  11 variables:
## $ PATTERN      : chr  "1    O F00" "1    I F00" "12   I F10" "12   O F10" ...
## $ PATTERNID    : int   198410 198414 198501 198498 198503 198508 198516 198514 1985
21 198520 ...
## $ ROUTE_NAME   : chr   "1" "1" "12" "12" ...
## $ DIRECTION    : chr   "O" "I" "I" "O" ...
## $ PATTERN_TYPE : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ SUB_TYPE     : int    0 0 1 1 0 0 1 1 0 0 ...
## $ PATTERN_VERSION: int    0 0 0 0 0 0 0 0 0 0 ...
## $ LINEABBR     : chr   "001" "001" "012" "012" ...
## $ SIGNID       : int   134 134 134 134 134 134 134 134 134 134 ...
## $ SERVICE_CA   : chr   "Frequent" "Frequent" "" "" ...
## $ shape        : chr   "MULTILINESTRING ((-122.396965 37.795437, -122.39781 37.7944
37, -122.39898 37.79429, -122.40016 37.79414, -122.4| __truncated__ "MULTILINESTRING
((-122.49315 37.77971, -122.49346 37.78163, -122.492386 37.78168, -122.49248 37.782917,
-122.49| __truncated__ "MULTILINESTRING ((-122.4192 37.748158, -122.42029 37.74813, -12
2.420364 37.748913, -122.41979 37.748947, -122.4| __truncated__ "MULTILINESTRING ((-12
2.42306 37.79394, -122.42324 37.794823, -122.42159 37.795033, -122.419945 37.795242, -12
2."| __truncated__ ...
```

```
SFMuni.stops <- read.csv("/Users/ellie/Downloads/Muni_Stops.csv")
str(SFMuni.stops)
```

```
## 'data.frame':    3249 obs. of  24 variables:
## $ OBJECTID      : int  36073 33704 33872 33441 35962 36243 34815 34726
35673 35126 ...
## $ STOPNAME      : chr   "Powell St&Francisco St SE-NS/BZ" "Chestnut St&
Laguna St SW-NS/BZ" "Geary Blvd&Fillmore St NW-FS/BZ" "3rd St&Folsom St N-FS/BZ" ...
## $ TRAPEZESTOPABBR : chr   "POWLFNC1" "CHESLGN1" "GEARFIL0" ".3STFOL0" ...
## $ RUCUSSTOPABBR  : chr   "POWLFNCO" "CHESLGNA" "GEARFILL" "3STFOLS" ...
## $ STOPID        : int   6056 3948 4295 3124 6098 7624 4878 5404 4481 70
38 ...
## $ LATITUDE       : num   37.8 37.8 37.8 37.8 37.8 ...
## $ LONGITUDE      : num  -122 -122 -122 -122 -122 ...
## $ ACCESSIBILITYMASK : int   0 0 0 0 0 0 0 0 0 0 ...
## $ ATSTREET       : chr   "FRANCISCO ST" "LAGUNA ST" "AVERY ST" "CLEMENTI
NA ST" ...
## $ ONSTREET       : chr   "POWELL ST" "CHESTNUT ST" "GEARY BLVD" "03RD S
T" ...
## $ POSITION        : chr   "NS" "NS" "FS" "FS" ...
## $ ORIENTATION    : chr   "SE" "SW" "NW" "NO" ...
## $ SERVICEPLANNINGSTOPTYPE : chr   "BZ" "BZ" "BZ" "BZ" ...
## $ SHELTER        : int   1 0 1 1 0 0 0 1 0 0 ...
## $ INSERT_TIMESTAMP : num   2.02e+13 2.02e+13 2.02e+13 2.02e+13 2.02e+13
...
## $ SDE_ID         : int   14257412 14253217 14253429 14252566 14258040 14
258356 14255435 14254611 14257301 14255518 ...
## $ SIGNUPID       : int   134 134 134 134 134 134 134 134 134 134 ...
## $ SUPERVISOR_DISTRICT : logi   NA NA NA NA NA NA ...
## $ shape          : chr   "POINT (-122.41165 37.80481)" "POINT (-122.4314
04 37.80137)" "POINT (-122.43305 37.78439)" "POINT (-122.39932 37.784203)" ...
## $ Neighborhoods  : int   106 17 103 32 103 43 45 21 20 99 ...
## $ SF.Find.Neighborhoods : int   106 17 103 32 103 43 45 21 20 99 ...
## $ Current.Police.Districts : int   6 4 4 1 8 10 7 4 5 6 ...
## $ Current.Supervisor.Districts: int   3 6 11 10 6 8 8 11 10 3 ...
## $ Analysis.Neighborhoods : int   23 13 39 8 31 16 41 39 36 23 ...
```

```
freq <- read.csv("/Users/ellie/Downloads/Bus_Frequency - Sheet1.csv")
```

```
str(freq)
```

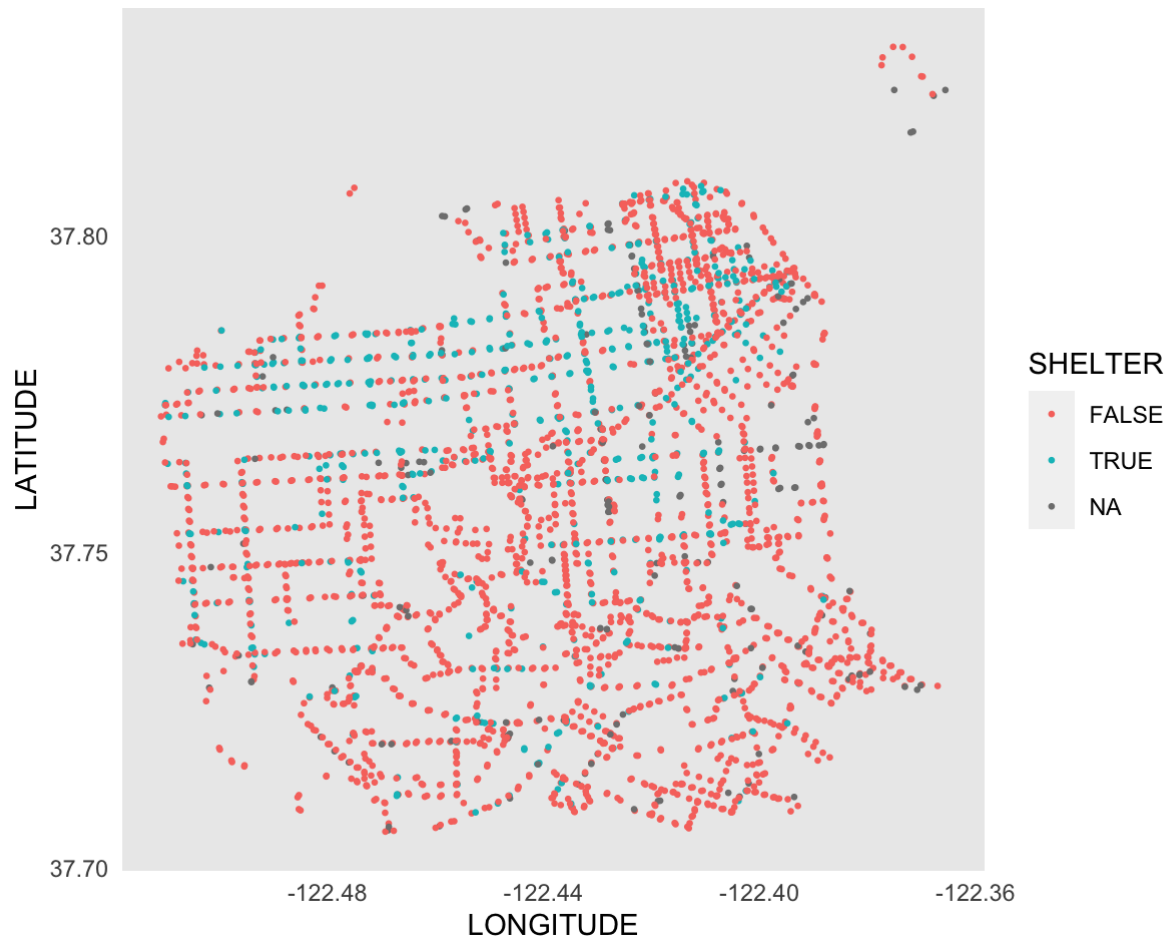
```
## 'data.frame':    54 obs. of  8 variables:
## $ Route          : chr   "PH Powell-Hyde Cable Car" "PM Powell-Mason Cable Car" "C
California Street Cable Car" "F Market & Wharves" ...
## $ Time.span      : chr   "7:00a - 11:00p" "7:00a - 11:00p" "7:00a - 9:00p" "7:00a -
10:00p" ...
## $ Morning        : chr   "17" "15" "17" "17" ...
## $ Midday         : chr   "10" "10" "10" "12" ...
## $ Evening        : chr   "13" "15" "12" "13" ...
## $ Late.Night     : chr   "17" "17" "20" "17" ...
## $ Owl            : chr   "-" "-" "-" "-" ...
## $ Average.Frequency: num   14.2 14.2 14.8 14.8 15 ...
```

```
SFMuni.stops$POSITION <- as.factor(SFMuni.stops$POSITION)
SFMuni.stops$ORIENTATION <- as.factor((SFMuni.stops$ORIENTATION))
SFMuni.stops$SHELTER <- as.logical(SFMuni.stops$SHELTER)
str(SFMuni.stops)
```

```
## 'data.frame':    3249 obs. of  24 variables:
##  $ OBJECTID          : int   36073 33704 33872 33441 35962 36243 34815 34726
35673 35126 ...
##  $ STOPNAME          : chr   "Powell St&Francisco St SE-NS/BZ" "Chestnut St&
Laguna St SW-NS/BZ" "Geary Blvd&Fillmore St NW-FS/BZ" "3rd St&Folsom St N-FS/BZ" ...
##  $ TRAPEZESTOPABBR   : chr   "POWLFNC1" "CHESLGN1" "GEARFIL0" ".3STFOL0" ...
##  $ RUCUSSTOPABBR     : chr   "POWLFNCO" "CHESLGNA" "GEARFILL" "3STFOLS" ...
##  $ STOPID            : int   6056 3948 4295 3124 6098 7624 4878 5404 4481 70
38 ...
##  $ LATITUDE          : num   37.8 37.8 37.8 37.8 37.8 ...
##  $ LONGITUDE         : num  -122 -122 -122 -122 -122 ...
##  $ ACCESSIBILITYMASK : int    0 0 0 0 0 0 0 0 0 0 ...
##  $ ATSTREET          : chr   "FRANCISCO ST" "LAGUNA ST" "AVERY ST" "CLEMENTI
NA ST" ...
##  $ ONSTREET          : chr   "POWELL ST" "CHESTNUT ST" "GEARY BLVD" "03RD S
T" ...
##  $ POSITION           : Factor w/ 9 levels "","FS","MB","MD",...: 7 7 2 2 2 3
3 2 2 7 ...
##  $ ORIENTATION       : Factor w/ 12 levels "","EA","MI","NE",...: 8 10 7 5 4
1 4 7 7 8 ...
##  $ SERVICEPLANNINGSTOPTYPE : chr   "BZ" "BZ" "BZ" "BZ" ...
##  $ SHELTER           : logi   TRUE FALSE TRUE TRUE FALSE FALSE ...
##  $ INSERT_TIMESTAMP   : num   2.02e+13 2.02e+13 2.02e+13 2.02e+13 2.02e+13
...
##  $ SDE_ID            : int   14257412 14253217 14253429 14252566 14258040 14
258356 14255435 14254611 14257301 14255518 ...
##  $ SIGNUPID          : int   134 134 134 134 134 134 134 134 134 134 ...
##  $ SUPERVISOR_DISTRICT : logi   NA NA NA NA NA NA ...
##  $ shape             : chr   "POINT (-122.41165 37.80481)" "POINT (-122.4314
04 37.80137)" "POINT (-122.43305 37.78439)" "POINT (-122.39932 37.784203)" ...
##  $ Neighborhoods     : int   106 17 103 32 103 43 45 21 20 99 ...
##  $ SF.Find.Neighborhoods : int   106 17 103 32 103 43 45 21 20 99 ...
##  $ Current.Police.Districts : int    6 4 4 1 8 10 7 4 5 6 ...
##  $ Current.Supervisor.Districts: int    3 6 11 10 6 8 8 11 10 3 ...
##  $ Analysis.Neighborhoods : int   23 13 39 8 31 16 41 39 36 23 ...
```

```
#scaled.LATITUDE <- SFMuni.stops$LATITUDE * 120
S1 <- ggplot(SFMuni.stops, aes(LONGITUDE, LATITUDE))

S1 + geom_point(size = 0.5, aes(color = SHELTER)) + theme(aspect.ratio=3/3, element_blan
k())
```



```
#s1 + geom_point(data = SFMuni.stops %>% filter(SHELTER == T), size = .7, aes(color = ORIENTATION, shape = SHELTER)) + theme(aspect.ratio=3/3)
#SFMuni.stops$SHELTER

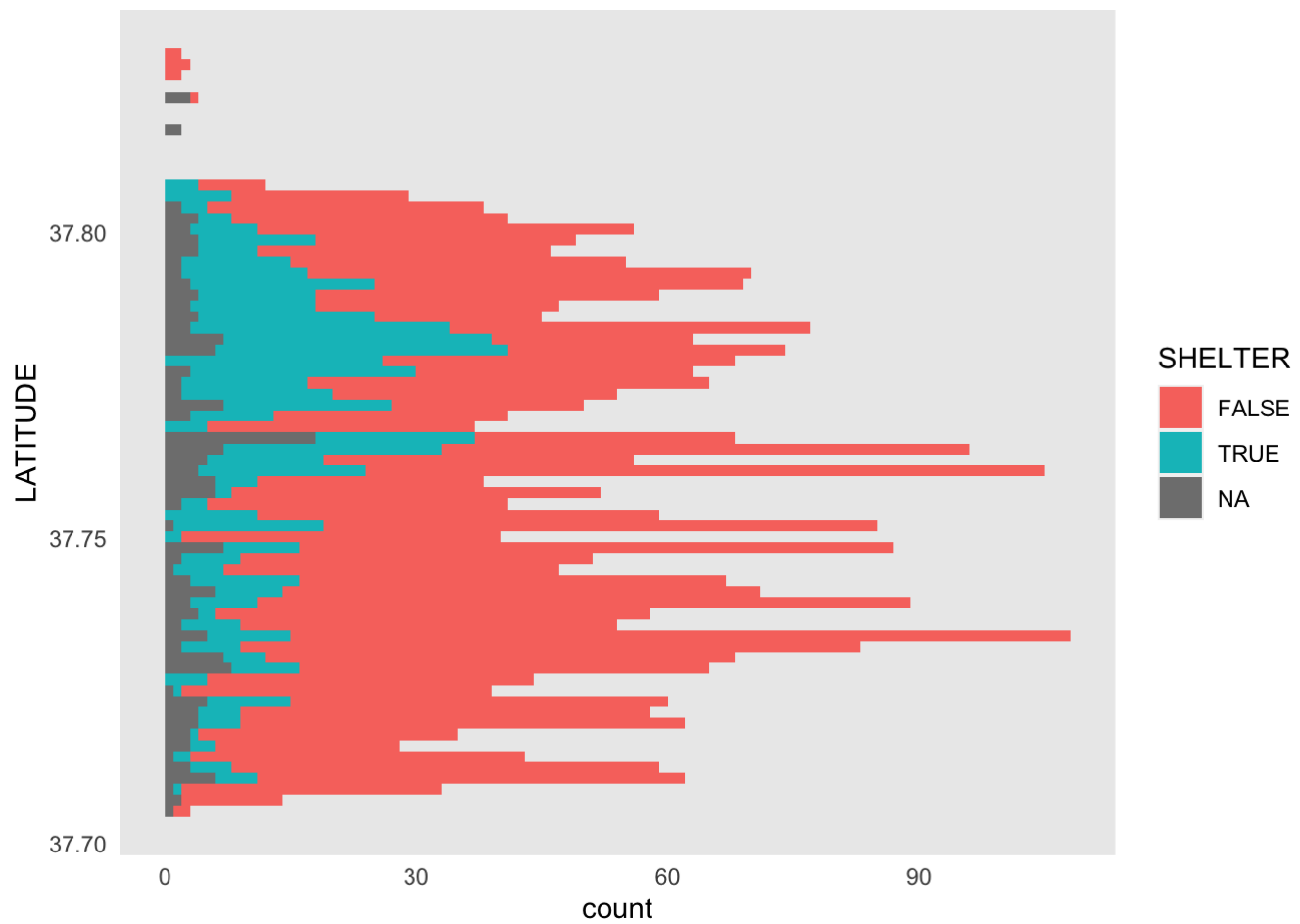
#s1 + geom_point(size = 1, aes(color = Analysis.Neighborhoods, shape = ORIENTATION)) + theme(aspect.ratio=3/3)
```

```
s1 <- ggplot(SFMuni.stops, aes(x=LONGITUDE))
#s1 + geom_histogram(aes(fill = SHELTER), bins = 50)

s11 <- ggplot(SFMuni.stops, aes(y=LATITUDE))
#p11 <- s11 + geom_histogram(aes(fill = SHELTER, alpha = .1), bins = 70)
#p11

#s12 <- ggplot(SFMuni.stops, aes(y=LATITUDE))
#p12 <- s12 + geom_histogram(aes(fill = SHELTER, alpha = .1), bins = 40)

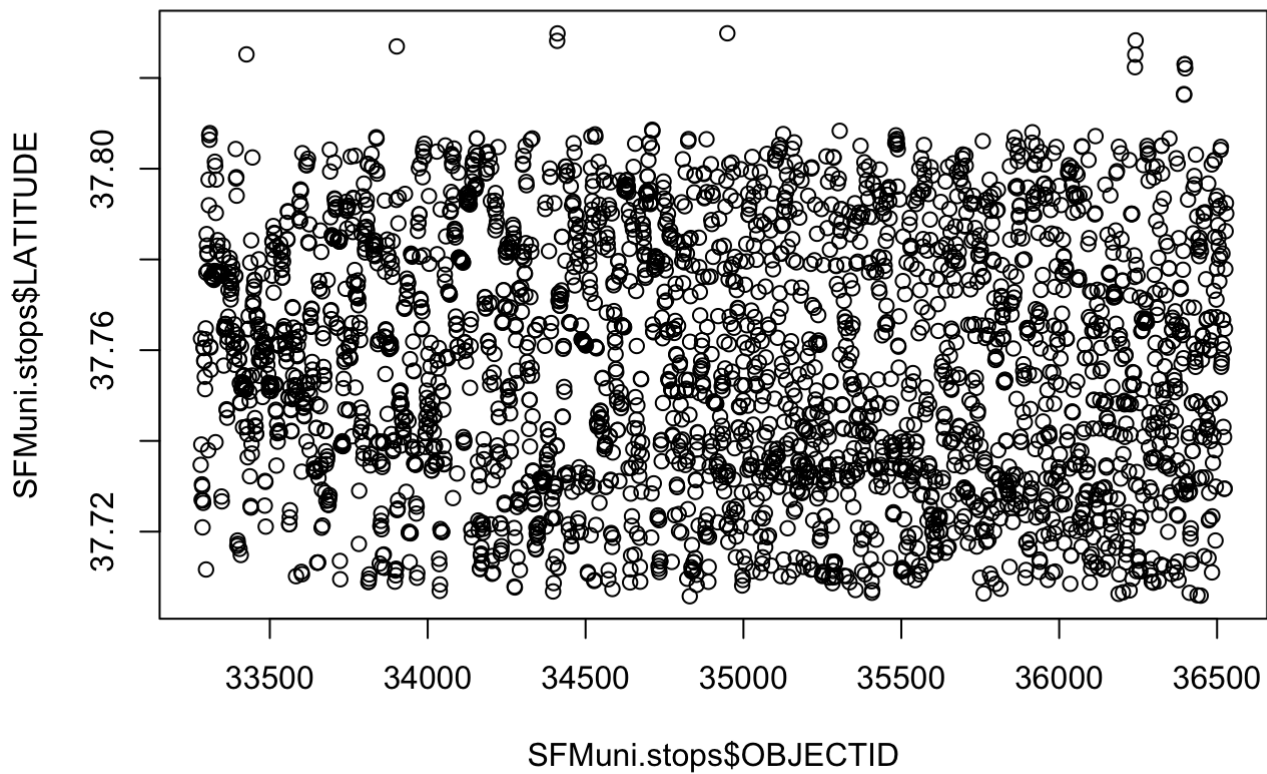
s11 + geom_histogram(aes(fill = SHELTER), bins = 70) + theme(element_blank()) #change bins back to 70
# + geom_histogram(aes(fill = SHELTER, alpha = .1), bins = 30) + geom_histogram(aes(fill = SHELTER, alpha = .1), bins = 10) + geom_histogram(aes(fill = SHELTER, alpha = .1), bins = 5)
```



```
#IDKplot <- ggplot(SFMuni.stops, x = ORIENTATION, y = POSITION)
#IDKplot + geom_curve()

for(y in 1:ncol(SFMuni.stops)){
  if (is.na(SFMuni.stops[1,y])) SFMuni.stops[1,y] = 0
}

plot(SFMuni.stops$OBJECTID, SFMuni.stops$LATITUDE)
```



```
#for (i in 1:nrow(SFMuni.stops)){
#  for (j in 1:nrow(SFMuni.stops)){
#    if (i == j){
#      cat("skipping", i, "and", j, "\n")
#    }else{
#      #l1 = SFMuni.stops$LONGITUDE[i]
#      #l2 = SFMuni.stops$LONGITUDE[j]
#      #a1 = SFMuni.stops$LATITUDE[i]
#      #a2 = SFMuni.stops$LATITUDE[j]
#      d1 = min()
#      #cat("we have a longitude of", l1, "and", l2, "\n")
#      #cat("we have a latitude of", a1, "and", a2, "\n")
#    }
#  }
# }
#}
```

```
#install.packages("sf")
#install.packages("leaflet")
library(dplyr)
library(sf)
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
library(leaflet)
#SFtransit$mapShape <- st_as_sfc(SFtransit$shape) # doesn't work for some reason
#plot34 <- ggplot(SFtransit, x = mapShape, y = shape) + geom_line(aes(x=mapShape,y=shape))
#plot34

str(SFtransit)
```

```
## 'data.frame':    139 obs. of  11 variables:
##  $ PATTERN      : chr  "1    O F00" "1    I F00" "12    I F10" "12    O F10" ...
##  $ PATTERNID     : int   198410 198414 198501 198498 198503 198508 198516 198514 1985
21 198520 ...
##  $ ROUTE_NAME    : chr   "1" "1" "12" "12" ...
##  $ DIRECTION     : chr   "O" "I" "I" "O" ...
##  $ PATTERN_TYPE  : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
##  $ SUB_TYPE      : int    0 0 1 1 0 0 1 1 0 0 ...
##  $ PATTERN_VERSION: int    0 0 0 0 0 0 0 0 0 0 ...
##  $ LINEABBR      : chr   "001" "001" "012" "012" ...
##  $ SIGNID        : int   134 134 134 134 134 134 134 134 134 134 ...
##  $ SERVICE_CA    : chr   "Frequent" "Frequent" "" "" ...
##  $ shape         : chr   "MULTILINESTRING ((-122.396965 37.795437, -122.39781 37.7944
37, -122.39898 37.79429, -122.40016 37.79414, -122.4"| __truncated__ "MULTILINESTRING
((-122.49315 37.77971, -122.49346 37.78163, -122.492386 37.78168, -122.49248 37.782917,
-122.49"| __truncated__ "MULTILINESTRING ((-122.4192 37.748158, -122.42029 37.74813, -12
2.420364 37.748913, -122.41979 37.748947, -122.4"| __truncated__ "MULTILINESTRING ((-12
2.42306 37.79394, -122.42324 37.794823, -122.42159 37.795033, -122.419945 37.795242, -12
2."| __truncated__ ...
```

```
#SFtransit$shape
```

```
l1shape <- as.numeric(SFtransit$shape[1])
```

```
## Warning: NAs introduced by coercion
```

```
l1shape
```

```
## [1] NA
```



```

SFtransit$shape <- gsub("MULTILINESTRING \\(\\(", "", SFtransit$shape)
SFtransit$shape <- gsub("\\)\\)", "", SFtransit$shape)

mapshape <- SFtransit$shape
longlang <- list()
lists <- list()
for(x in 1:length(mapshape[ ])){

  longlang <- strsplit(mapshape[x], " ,")
  longlang <- strsplit(mapshape[x], " +")
  sfshape <- longlang[[1]]
  sfshape <- gsub(",", "", sfshape)
  as.numeric(sfshape)

  x <- list()
  y <- list()
  for(i in 1:length(sfshape)){
    if(i%%2!=0){
      x <- append(x, sfshape[i])
    }else{
      y <- append(y, sfshape[i])
    }
  }
  xs <- vector()
  ys <- vector()
  for(i in 1:length(x[ ])){
    xs[i] <- x[[i]]
    ys[i] <- y[[i]]
  }
  df <- data.frame()
  df <-data.frame(x=c(xs),
                  y=c(ys))
  lists[[length(lists)+1]] <- df
}

```

```

#for (i in 1:length(lists)){
#  lists[[i]] <- as.numeric(lists[[i]])
#}

#####

#lists[[1]]$x

#str(lists)

mySPECIALdf <- as.data.frame(lists[[1]])
mySPECIALdf$x <- as.numeric(mySPECIALdf$x)
mySPECIALdf$y <- as.numeric(mySPECIALdf$y)
#str(mySPECIALdf)

for (i in 1:length(lists)){
  #print(i)
  lists[[i]]$x <- as.numeric(lists[[i]]$x)
  lists[[i]]$y <- as.numeric(lists[[i]]$y)
}

str(lists)

```

```

## List of 139
## $ : 'data.frame':   98 obs. of  2 variables:
## ..$ x: num [1:98] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:98] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   99 obs. of  2 variables:
## ..$ x: num [1:99] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:99] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  114 obs. of  2 variables:
## ..$ x: num [1:114] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:114] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':  112 obs. of  2 variables:
## ..$ x: num [1:112] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:112] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  147 obs. of  2 variables:
## ..$ x: num [1:147] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:147] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  149 obs. of  2 variables:
## ..$ x: num [1:149] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:149] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':  168 obs. of  2 variables:
## ..$ x: num [1:168] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:168] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':  162 obs. of  2 variables:
## ..$ x: num [1:162] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:162] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  156 obs. of  2 variables:
## ..$ x: num [1:156] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:156] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':  180 obs. of  2 variables:
## ..$ x: num [1:180] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:180] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  149 obs. of  2 variables:
## ..$ x: num [1:149] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:149] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':  152 obs. of  2 variables:
## ..$ x: num [1:152] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:152] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  123 obs. of  2 variables:
## ..$ x: num [1:123] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:123] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  151 obs. of  2 variables:
## ..$ x: num [1:151] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:151] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   99 obs. of  2 variables:
## ..$ x: num [1:99] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:99] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  110 obs. of  2 variables:
## ..$ x: num [1:110] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:110] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':  214 obs. of  2 variables:
## ..$ x: num [1:214] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:214] 37.7 37.7 37.7 37.7 37.7 ...

```

```

## $ : 'data.frame':   191 obs. of  2 variables:
## ..$ x: num [1:191] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:191] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   139 obs. of  2 variables:
## ..$ x: num [1:139] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:139] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   147 obs. of  2 variables:
## ..$ x: num [1:147] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:147] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':    96 obs. of  2 variables:
## ..$ x: num [1:96] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:96] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   107 obs. of  2 variables:
## ..$ x: num [1:107] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:107] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   100 obs. of  2 variables:
## ..$ x: num [1:100] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:100] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    88 obs. of  2 variables:
## ..$ x: num [1:88] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:88] 37.8 37.8 37.8 37.8 37.7 ...
## $ : 'data.frame':   235 obs. of  2 variables:
## ..$ x: num [1:235] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:235] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   223 obs. of  2 variables:
## ..$ x: num [1:223] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:223] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   234 obs. of  2 variables:
## ..$ x: num [1:234] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:234] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   248 obs. of  2 variables:
## ..$ x: num [1:248] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:248] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   274 obs. of  2 variables:
## ..$ x: num [1:274] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:274] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   286 obs. of  2 variables:
## ..$ x: num [1:286] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:286] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    71 obs. of  2 variables:
## ..$ x: num [1:71] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:71] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   102 obs. of  2 variables:
## ..$ x: num [1:102] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:102] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    97 obs. of  2 variables:
## ..$ x: num [1:97] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:97] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    86 obs. of  2 variables:
## ..$ x: num [1:86] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:86] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   114 obs. of  2 variables:

```

```
## ..$ x: num [1:114] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:114] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 100 obs. of 2 variables:
## ..$ x: num [1:100] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:100] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 116 obs. of 2 variables:
## ..$ x: num [1:116] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:116] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 97 obs. of 2 variables:
## ..$ x: num [1:97] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:97] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 132 obs. of 2 variables:
## ..$ x: num [1:132] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:132] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 127 obs. of 2 variables:
## ..$ x: num [1:127] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:127] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 74 obs. of 2 variables:
## ..$ x: num [1:74] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:74] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 86 obs. of 2 variables:
## ..$ x: num [1:86] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:86] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 173 obs. of 2 variables:
## ..$ x: num [1:173] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:173] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 302 obs. of 2 variables:
## ..$ x: num [1:302] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:302] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 199 obs. of 2 variables:
## ..$ x: num [1:199] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:199] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 180 obs. of 2 variables:
## ..$ x: num [1:180] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:180] 37.8 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 110 obs. of 2 variables:
## ..$ x: num [1:110] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:110] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 111 obs. of 2 variables:
## ..$ x: num [1:111] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:111] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 155 obs. of 2 variables:
## ..$ x: num [1:155] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:155] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 151 obs. of 2 variables:
## ..$ x: num [1:151] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:151] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 149 obs. of 2 variables:
## ..$ x: num [1:149] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:149] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 117 obs. of 2 variables:
## ..$ x: num [1:117] -123 -123 -123 -123 -123 ...
```

```
## ..$ y: num [1:117] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 117 obs. of 2 variables:
## ..$ x: num [1:117] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:117] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 111 obs. of 2 variables:
## ..$ x: num [1:111] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:111] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 70 obs. of 2 variables:
## ..$ x: num [1:70] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:70] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 77 obs. of 2 variables:
## ..$ x: num [1:77] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:77] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 334 obs. of 2 variables:
## ..$ x: num [1:334] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:334] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 336 obs. of 2 variables:
## ..$ x: num [1:336] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:336] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 299 obs. of 2 variables:
## ..$ x: num [1:299] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:299] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 301 obs. of 2 variables:
## ..$ x: num [1:301] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:301] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 72 obs. of 2 variables:
## ..$ x: num [1:72] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:72] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 80 obs. of 2 variables:
## ..$ x: num [1:80] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:80] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 181 obs. of 2 variables:
## ..$ x: num [1:181] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:181] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 184 obs. of 2 variables:
## ..$ x: num [1:184] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:184] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 141 obs. of 2 variables:
## ..$ x: num [1:141] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:141] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 136 obs. of 2 variables:
## ..$ x: num [1:136] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:136] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 108 obs. of 2 variables:
## ..$ x: num [1:108] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:108] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 107 obs. of 2 variables:
## ..$ x: num [1:107] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:107] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 127 obs. of 2 variables:
## ..$ x: num [1:127] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:127] 37.7 37.7 37.7 37.7 37.7 ...
```

```

## $ : 'data.frame':   139 obs. of  2 variables:
## ..$ x: num [1:139] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:139] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   292 obs. of  2 variables:
## ..$ x: num [1:292] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:292] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   120 obs. of  2 variables:
## ..$ x: num [1:120] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:120] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   165 obs. of  2 variables:
## ..$ x: num [1:165] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:165] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   247 obs. of  2 variables:
## ..$ x: num [1:247] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:247] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':    52 obs. of  2 variables:
## ..$ x: num [1:52] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:52] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    39 obs. of  2 variables:
## ..$ x: num [1:39] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:39] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    86 obs. of  2 variables:
## ..$ x: num [1:86] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:86] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':    75 obs. of  2 variables:
## ..$ x: num [1:75] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:75] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   131 obs. of  2 variables:
## ..$ x: num [1:131] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:131] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   112 obs. of  2 variables:
## ..$ x: num [1:112] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:112] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   173 obs. of  2 variables:
## ..$ x: num [1:173] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:173] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   154 obs. of  2 variables:
## ..$ x: num [1:154] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:154] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame':   106 obs. of  2 variables:
## ..$ x: num [1:106] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:106] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   107 obs. of  2 variables:
## ..$ x: num [1:107] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:107] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':   107 obs. of  2 variables:
## ..$ x: num [1:107] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:107] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    58 obs. of  2 variables:
## ..$ x: num [1:58] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:58] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame':    54 obs. of  2 variables:

```

```
## ..$ x: num [1:54] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:54] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 54 obs. of 2 variables:
## ..$ x: num [1:54] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:54] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 81 obs. of 2 variables:
## ..$ x: num [1:81] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:81] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 124 obs. of 2 variables:
## ..$ x: num [1:124] -123 -123 -123 -123 -123 ...
## ..$ y: num [1:124] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 125 obs. of 2 variables:
## ..$ x: num [1:125] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:125] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 161 obs. of 2 variables:
## ..$ x: num [1:161] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:161] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 160 obs. of 2 variables:
## ..$ x: num [1:160] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:160] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 265 obs. of 2 variables:
## ..$ x: num [1:265] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:265] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 258 obs. of 2 variables:
## ..$ x: num [1:258] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:258] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 178 obs. of 2 variables:
## ..$ x: num [1:178] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:178] 37.8 37.8 37.8 37.8 37.8 ...
## $ : 'data.frame': 186 obs. of 2 variables:
## ..$ x: num [1:186] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:186] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 222 obs. of 2 variables:
## ..$ x: num [1:222] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:222] 37.7 37.7 37.7 37.7 37.7 ...
## $ : 'data.frame': 254 obs. of 2 variables:
## ..$ x: num [1:254] -122 -122 -122 -122 -122 ...
## ..$ y: num [1:254] 37.8 37.8 37.8 37.8 37.8 ...
## [list output truncated]
```

```
max(SFMuni.stops$LATITUDE)
```

```
## [1] 37.82985
```

```
min(SFMuni.stops$LATITUDE)
```

```
## [1] 37.70576
```



```
max(SFMuni.stops$LONGITUDE)
```

```
## [1] -122.3656
```

```
min(SFMuni.stops$LONGITUDE)
```

```
## [1] -122.5108
```

```
first <- as.data.frame(lists[[2]])  
pl <- ggplot(first, aes(x,y))  
pl1 <- pl + geom_path(size = .3) + xlim(-122.5108,-122.3656) + ylim(37.70576,37.82985) +  
theme(aspect.ratio = 1/1)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## i Please use `linewidth` instead.
```

```
pl1 + geom_point(data = SFMuni.stops, size = 0.5, aes(LONGITUDE, LATITUDE, color = SHELTER)) + theme(aspect.ratio=3/3, element_blank())
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



```
lists[[23]]
```

##	x	y
## 1	-122.4233	37.79396
## 2	-122.4249	37.79462
## 3	-122.4232	37.79483
## 4	-122.4232	37.79482
## 5	-122.4216	37.79503
## 6	-122.4214	37.79415
## 7	-122.4212	37.79326
## 8	-122.4196	37.79347
## 9	-122.4179	37.79369
## 10	-122.4178	37.79281
## 11	-122.4177	37.79234
## 12	-122.4176	37.79193
## 13	-122.4174	37.79098
## 14	-122.4172	37.79005
## 15	-122.4170	37.78912
## 16	-122.4154	37.78933
## 17	-122.4137	37.78954
## 18	-122.4136	37.78860
## 19	-122.4134	37.78796
## 20	-122.4134	37.78767
## 21	-122.4133	37.78739
## 22	-122.4132	37.78674
## 23	-122.4130	37.78579
## 24	-122.4129	37.78536
## 25	-122.4129	37.78531
## 26	-122.4128	37.78487
## 27	-122.4126	37.78394
## 28	-122.4124	37.78301
## 29	-122.4122	37.78208
## 30	-122.4120	37.78114
## 31	-122.4132	37.78100
## 32	-122.4137	37.78093
## 33	-122.4153	37.78072
## 34	-122.4151	37.77979
## 35	-122.4150	37.77889
## 36	-122.4148	37.77873
## 37	-122.4142	37.77824
## 38	-122.4132	37.77746
## 39	-122.4126	37.77703
## 40	-122.4122	37.77667
## 41	-122.4116	37.77623
## 42	-122.4111	37.77580
## 43	-122.4106	37.77543
## 44	-122.4101	37.77500
## 45	-122.4095	37.77455
## 46	-122.4091	37.77420
## 47	-122.4085	37.77377
## 48	-122.4097	37.77284
## 49	-122.4101	37.77253
## 50	-122.4107	37.77206
## 51	-122.4113	37.77158

52 -122.4125 37.77063
53 -122.4140 37.77187
54 -122.4145 37.77141
55 -122.4150 37.77090
56 -122.4153 37.77036
57 -122.4156 37.76960
58 -122.4156 37.76937
59 -122.4156 37.76854
60 -122.4155 37.76694
61 -122.4153 37.76532
62 -122.4131 37.76545
63 -122.4124 37.76550
64 -122.4115 37.76555
65 -122.4105 37.76561
66 -122.4104 37.76432
67 -122.4103 37.76304
68 -122.4101 37.76176
69 -122.4100 37.76049
70 -122.4099 37.75957
71 -122.4098 37.75920
72 -122.4097 37.75911
73 -122.4097 37.75845
74 -122.4096 37.75773
75 -122.4096 37.75760
76 -122.4095 37.75609
77 -122.4095 37.75600
78 -122.4094 37.75497
79 -122.4093 37.75440
80 -122.4093 37.75426
81 -122.4093 37.75423
82 -122.4091 37.75280
83 -122.4090 37.75120
84 -122.4088 37.74960
85 -122.4087 37.74836
86 -122.4097 37.74834
87 -122.4106 37.74832
88 -122.4115 37.74830
89 -122.4126 37.74828
90 -122.4137 37.74826
91 -122.4149 37.74823
92 -122.4159 37.74822
93 -122.4160 37.74918
94 -122.4161 37.75077
95 -122.4163 37.75238
96 -122.4168 37.75234
97 -122.4174 37.75231
98 -122.4179 37.75228
99 -122.4185 37.75224
100 -122.4186 37.75229

```
#first <- as.data.frame(lists[[2]])
sl <- ggplot(SFMuni.stops, aes(LONGITUDE, LATITUDE))
psl <- sl + geom_point(size = 0.45, aes(LONGITUDE, LATITUDE, color = SHELTER)) + theme(aspect.ratio=3/3, element_blank())
#psl
#psl + geom_path(data = lists[[2]], size = 0.3, aes(x,y)) + theme(aspect.ratio=3/3, element_blank())
```

```
#str(lists)
for (i in 1:length(lists)){
  newRoute = paste("geom_path(data = lists[[",i,"]], size = 0.3, aes(x,y, size = 12))",
  sep = "")
  psl <- psl + eval(parse(text = newRoute))

  #print(newRoute)
}
psl
```



```
?leaflet
?sf
```

Components of the write-up and presentation

Statement of question/topic that you want to answer and what motivated you to study the question/topic

The topic we studied was public transit in San Francisco. Specifically, we were interested in how the population density in the city relates to the MUNI routes. We all ride MUNI almost every day and were excited to learn more about the system and think through how potential improvements could be made.

If applicable, what data will you analyze? Identify at least one data source (can be from Kaggle, etc.)

We used two data sets. We downloaded both from sf.gov as .csv files. The first, "MUNI Stops", lists all 3428 MUNI bus stops in the city, including their cross sections, stop ID numbers, and latitudinal and longitudinal information. The second, "MUNI Routes" lists all of the bus lines and their information regarding their direction and frequency.

What challenges do you face in analyzing this data?

There were many challenges we faced in this process. First, it was very difficult to figure out how to get the data to a place where it could be plotted. We wanted to be able to plot the actual bus routes by graphing each stop onto a map of the city. The latitude/longitude data regarding each stop was not in a format that this was possible, so we had to first figure out how to list it in an accessible way. Second, despite our best efforts, we could not find a data set with information regarding the population density within the city. Our solution to this was to use a photo of a map of the city with the density information represented by different colors. We used this image to overlay our routes onto. This solution had its own problems associated with it, including scaling our map of the stops to the scale of the map used online since we did not have the ability to manipulate that map.

What packages were needed for this case study?

1. `dplyr` : Is used to organize data and make manipulation easier, allows one to call row/column data easily
2. `sf` : Is used to plot spatial data, experimented with this method to plot routes using polygon geometry
3. `leaflet` : Creates a map widget using the leaflet package, easily renders spatial objects from the `sp` or `sf` packages, or data frames with latitude/longitude columns
4. `ggplot` : Can be used to declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all layers
5. `ggpubr` : Simpler version of `ggplot`, creates plots with less code

Discussion: What did you learn from this experience? What more could you do with this project in the future?

One of the biggest lessons we learned is that data cleaning and organizing can be the most time consuming part of analysis projects. Although it took us a while to find the data we wanted to work with, develop questions to explore, and eventually to discuss the results, the most challenging part for us was to get the data into a format that we could work with. Part of the reason this was so time consuming was because of how important it was for us to set the data aside and come back to it multiple times with fresh(ish) eyes. On that note, another key part of our project was the teamwork. It took all of us sharing ideas, explaining our problems, and working together to get the assignment done. We were all happy to work in a collaborative and patient team and now have a much better understanding as to the value of teams in data analysis. Finally, there was so much more we wanted to do with this data if we had more time and more data to work with. Our ideas for additional analyses included examining the frequency information in relation to which stops had shelters and which did not, and looking at where car crashes occur most frequently in order to provide recommendations for better traffic signals and alternative MUNI routes. These analyses along with the analyses we completed in this project could be used to recommend changes to the frequency of bus routes, or even to provide recommendations for how to expand the underground transit options in the city.