Drive Time Analysis

Dusty and JD

6/20/2019

Read in and Clean Data

```r
library(tidyverse)
library(lubridate)
data = read_csv("drive_times.csv", col_names = FALSE) %>%
  rename(Home=X1,Work=X2,Date=X3,DriveTime=X4) %>% ## Make it easier to read
  filter(substr(Work,1,4)=="6003") %>% ## Place the first four letters of your "Work" Address here
  filter(substr(Home,1,4)=="8109") %>% ## Place the first four letters of your "Home" Address here
  select(-c(X5,X6)) ## Removes these columns for now

cleaned =
data %>%
  select(-c(Home,Work)) %>% ## temporarily removed for viewing
  mutate(Date = as.POSIXct(Date)) %>% ## must do this to make future comands work
  mutate(Date = with_tz(Date, tzone = "America/New_York")) %>% ## make this the eastern time zone
  mutate(DayOfWeek = weekdays(Date)) %>% ## find what day of week
  filter(DayOfWeek != "Sunday" && DayOfWeek != "Saturday") %>% ## filter out weekends
  mutate(Breaks = round_date(Date, "15 mins")) %>% ## floor to the lowest 15 minutes
  mutate(Breaks = as.POSIXct(cut(Breaks, breaks = "1 hour"))) %>% ## bin 1 hour -- eventually will do 1
  mutate(Breaks = as.factor(format(Breaks, "%H:%M:%S"))) ## extract out just the HMS for plotting purpo

cleaned
```

```
## # A tibble: 315 x 4
##    Date                DriveTime DayOfWeek Breaks
##    <dttm>                  <dbl> <chr>     <fct>
##  1 2019-06-17 15:07:14      1401 Monday    15:00:00
##  2 2019-06-17 15:17:30      1424 Monday    15:00:00
##  3 2019-06-17 16:45:12      1409 Monday    16:00:00
##  4 2019-06-17 17:01:30      1377 Monday    17:00:00
##  5 2019-06-17 17:01:45      1386 Monday    17:00:00
##  6 2019-06-17 17:01:53      1384 Monday    17:00:00
##  7 2019-06-17 17:02:47      1391 Monday    17:00:00
##  8 2019-06-17 17:17:50      1371 Monday    17:00:00
##  9 2019-06-17 17:32:53      1450 Monday    17:00:00
## 10 2019-06-17 17:47:56      1379 Monday    17:00:00
## # ... with 305 more rows
```
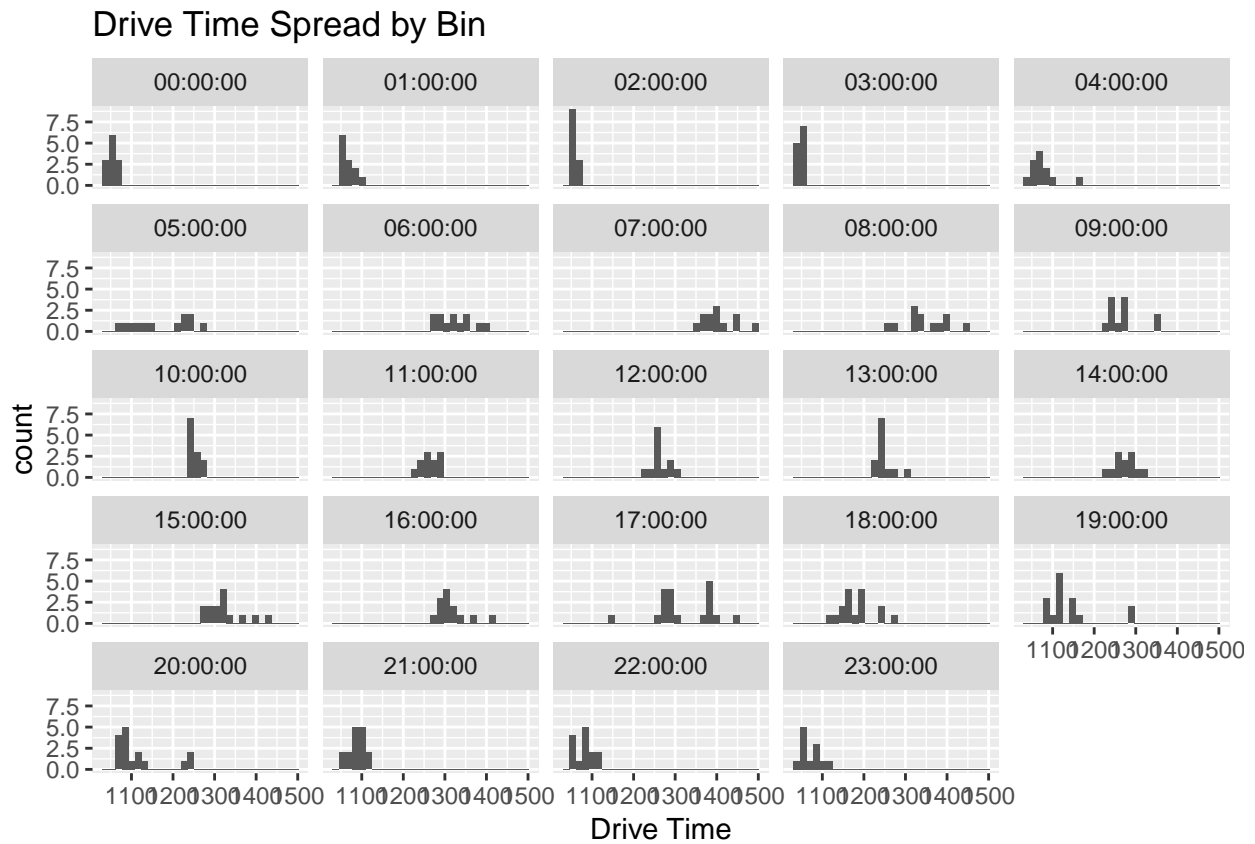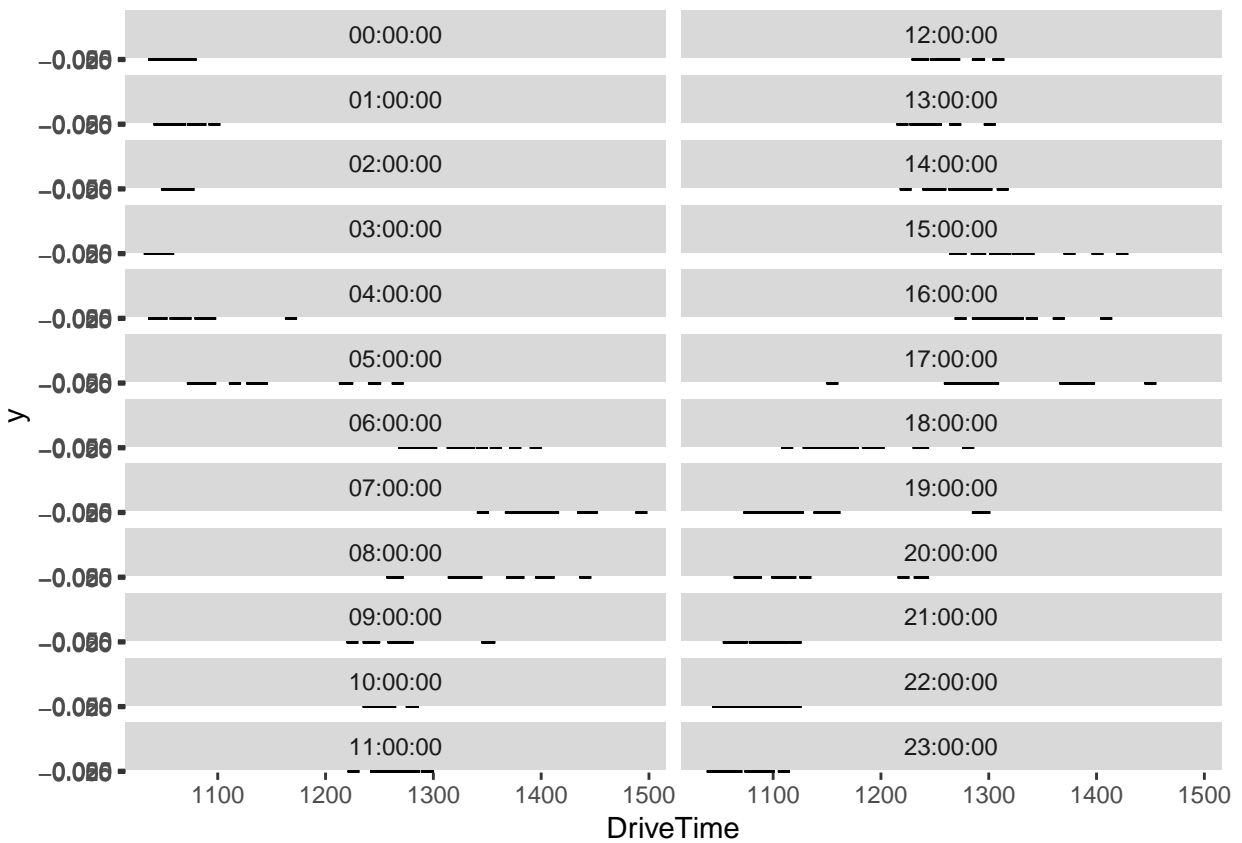
Histogram by Bin

```r
cleaned %>%
  ggplot(aes(x=DriveTime)) +
  geom_histogram() +
```

```
facet_wrap(~Breaks) +
labs(x="Drive Time", title = "Drive Time Spread by Bin")
```
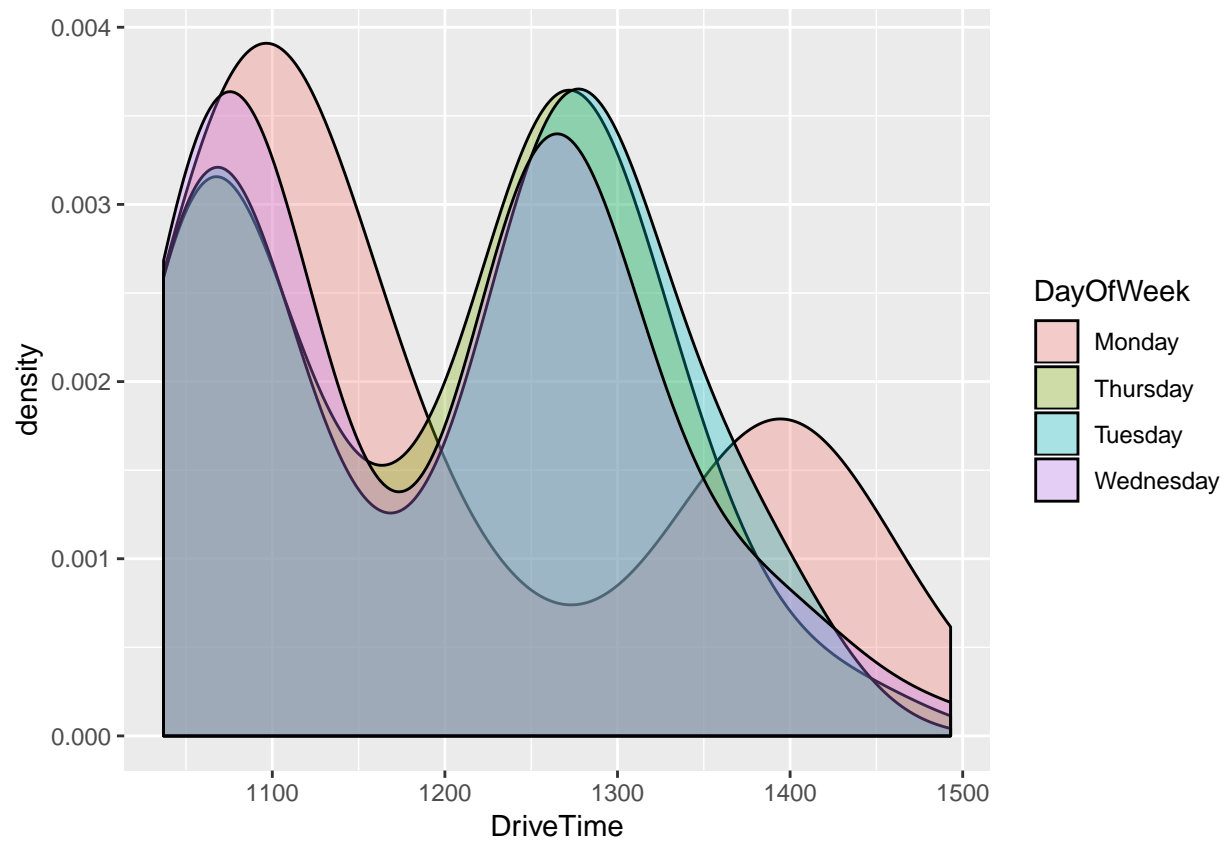

Drive Time Spread by Bin

Plot of time to see spread

```
cleaned %>%
  ggplot(aes(x=DriveTime,y=0)) +
  geom_point() +
  facet_wrap(~Breaks, ncol = 2,dir = "v")
```

Day of Week

```r
cleaned %>%
  ggplot(aes(x=DriveTime)) +
  geom_density(aes(fill = DayOfWeek), alpha = .3)
```

```
# theme(legend.position = "none")
```

Summary Statistics

```
cleaned %>%
  count(Breaks)
```

```
## # A tibble: 24 x 2
##    Breaks       n
##    <fct>    <int>
##  1 00:00:00    12
##  2 01:00:00    12
##  3 02:00:00    12
##  4 03:00:00    12
##  5 04:00:00    12
##  6 05:00:00    12
##  7 06:00:00    12
##  8 07:00:00    12
##  9 08:00:00    12
## 10 09:00:00    12
## # ... with 14 more rows
```

```r
cleaned %>%
  group_by(Breaks) %>%
  summarise(mean = mean(DriveTime), sd = sd(DriveTime))
```

```
## # A tibble: 24 x 3
##    Breaks     mean    sd
##    <fct>     <dbl> <dbl>
##  1 00:00:00 1056. 10.4
##  2 01:00:00 1066. 15.2
##  3 02:00:00 1059.  5.66
##  4 03:00:00 1046.  4.92
##  5 04:00:00 1076. 32.8
##  6 05:00:00 1172. 70.4
##  7 06:00:00 1322. 40.2
##  8 07:00:00 1404. 39.8
##  9 08:00:00 1347. 54.6
## 10 09:00:00 1271. 40.8
## # ... with 14 more rows
```