Does a change in scenery cause a change in lifestyle?
Stat231: Google Calendar Report

Mike Santos

Due Friday, September 25 by 2:00 PM PDT

How do I spend my time?

```
#Setting up table and altering column titles
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
Log <- read_csv("Log.csv")
```

```
## Parsed with column specification:
## cols(
##   START.DATE = col_datetime(format = ""),
##   END.DATE = col_datetime(format = ""),
##   START.TIME.LOCAL = col_character(),
##   END.TIME.LOCAL = col_character(),
##   DURATION = col_double(),
##   MINUTES = col_double(),
##   NAME = col_character(),
##   LOCATION = col_character()
## )
```

```
log = Log %>% filter (Log$START.TIME.LOCAL > "2020-08-15")
log$general.location = "Road Trip"
log$time.zone = "CDT"
Eastern = "EDT"
Central = "CDT"
```

```
Mountain = "MDT"
Pacific = "PDT"

#Adding time zone and location to table
logEast = log %>% filter (grepl(Eastern, log$START.TIME.LOCAL))
logEast$time.zone = "EDT"
logCentral =  log %>% filter (grepl(Central, log$START.TIME.LOCAL))
logCentral$time.zone = "CDT"
logRT = rbind(logEast, logCentral)
logDEN = log %>% filter (grepl(Mountain, log$START.TIME.LOCAL))
logDEN$time.zone = "MDT"
logLA = log %>% filter (grepl(Pacific, log$START.TIME.LOCAL))
logLA$time.zone = "PDT"

#Adding general location & combining tables
logRT$general.location = "Road Trip"
logDEN$general.location = "Denver"
logLA$general.location = "Los Angeles"
time.log = rbind (logRT, logDEN, logLA)
time.log = time.log %>% rename(activity = "NAME")
time.log = time.log %>% rename(time.seconds = "DURATION")
time.log = time.log %>% rename(time.minutes = "MINUTES")
```

Describe your question(s) here. Briefly describe your data collection process, including how you defined variables of interest.

My questions are all regarding the differences in my behavior and how I have spent my time differently throughout my travels which I have categorized into three different sections of time (Road Trip, Denver, Los Angeles). Using an app that tracks changes in location, I am trying to notice patterns or blatant differences in lifestyle.

Describe what information is conveyed through data visualization #1 (that you'll create below) here.

```
library(ggplot2)
print("This visualization showcases my average time spent on activities in each location. While this is
```

```
## [1] "This visualization showcases my average time spent on activities in each location. While this is
```

```
#Calculating average time spent per activity for each location.
print("Road trip:")
```

```
## [1] "Road trip:"
```

```
RT.time = tapply(logRT$MINUTES, logRT$NAME, mean)
RT.time
```

```
##   Chillin     Cycle    Flight      Food      Rest     Sleep Transport     Trips
## 104.18333  23.30000 139.38333  45.85556  18.96667 422.94444  56.31053  21.01667
##      Walk      Work
##  15.66818  23.85000
```

```r
print("Denver:")
```

```
## [1] "Denver:"
```

```r
DEN.time = tapply(logDEN$MINUTES, logDEN$NAME, mean)
DEN.time
```

```
##    Chillin     Flight       Food     Hiking   Shopping      Sleep  Transport
## 1018.35595  126.03333   79.01111   29.30833   77.40000  379.96528   41.39085
##      Trips       Walk
##   26.02500   17.83080
```

```r
print("Los Angeles:")
```

```
## [1] "Los Angeles:"
```

```r
LA.time = tapply(logLA$MINUTES, logLA$NAME, mean)
LA.time
```

```
##      Beach    Chillin     Flight       Food   Shopping      Sleep  Transport      Trips
## 194.71667  475.61296  172.41667   58.66429   29.90000  634.09226   23.37793   43.98333
##       Walk
##   13.63636
```

```r
#Changing from array to data frame, adding location.
RT.time = as.data.frame.table(RT.time, col.names=pop)
RT.time$Location = "Road Trip"

DEN.time = as.data.frame.table(DEN.time, col.names=pop)
DEN.time$Location = "Denver"

LA.time = as.data.frame.table(LA.time, col.names=pop)
LA.time$Location = "Los Angeles"

#Combining 3 separate data frames into one
time.table = rbind(RT.time, DEN.time, LA.time)
names(time.table) = c("Activity", "Average.Time.Spent", "Location")
time.table
```

```
##      Activity Average.Time.Spent   Location
## 1     Chillin          104.18333   Road Trip
## 2       Cycle           23.30000   Road Trip
## 3      Flight          139.38333   Road Trip
## 4        Food           45.85556   Road Trip
## 5        Rest           18.96667   Road Trip
## 6       Sleep          422.94444   Road Trip
## 7   Transport           56.31053   Road Trip
## 8       Trips           21.01667   Road Trip
## 9        Walk           15.66818   Road Trip
## 10       Work           23.85000   Road Trip
```

```
## 11    Chillin      1018.35595      Denver
## 12     Flight       126.03333      Denver
## 13       Food        79.01111      Denver
## 14     Hiking        29.30833      Denver
## 15   Shopping        77.40000      Denver
## 16      Sleep       379.96528      Denver
## 17  Transport        41.39085      Denver
## 18      Trips        26.02500      Denver
## 19       Walk        17.83080      Denver
## 20      Beach       194.71667 Los Angeles
## 21    Chillin       475.61296 Los Angeles
## 22     Flight       172.41667 Los Angeles
## 23       Food        58.66429 Los Angeles
## 24   Shopping        29.90000 Los Angeles
## 25      Sleep       634.09226 Los Angeles
## 26  Transport        23.37793 Los Angeles
## 27      Trips        43.98333 Los Angeles
## 28       Walk        13.63636 Los Angeles
```

```
#Creating bar graph, average time spent per activity
plot = ggplot(time.table, aes(fill=Location, x=Activity, y=Average.Time.Spent)) + geom_bar(stat="identi
  labs(x="Activity", y="Average Time Spent (Minutes)", title = "Graph 1.1 - Average time spent per acti
plot
```
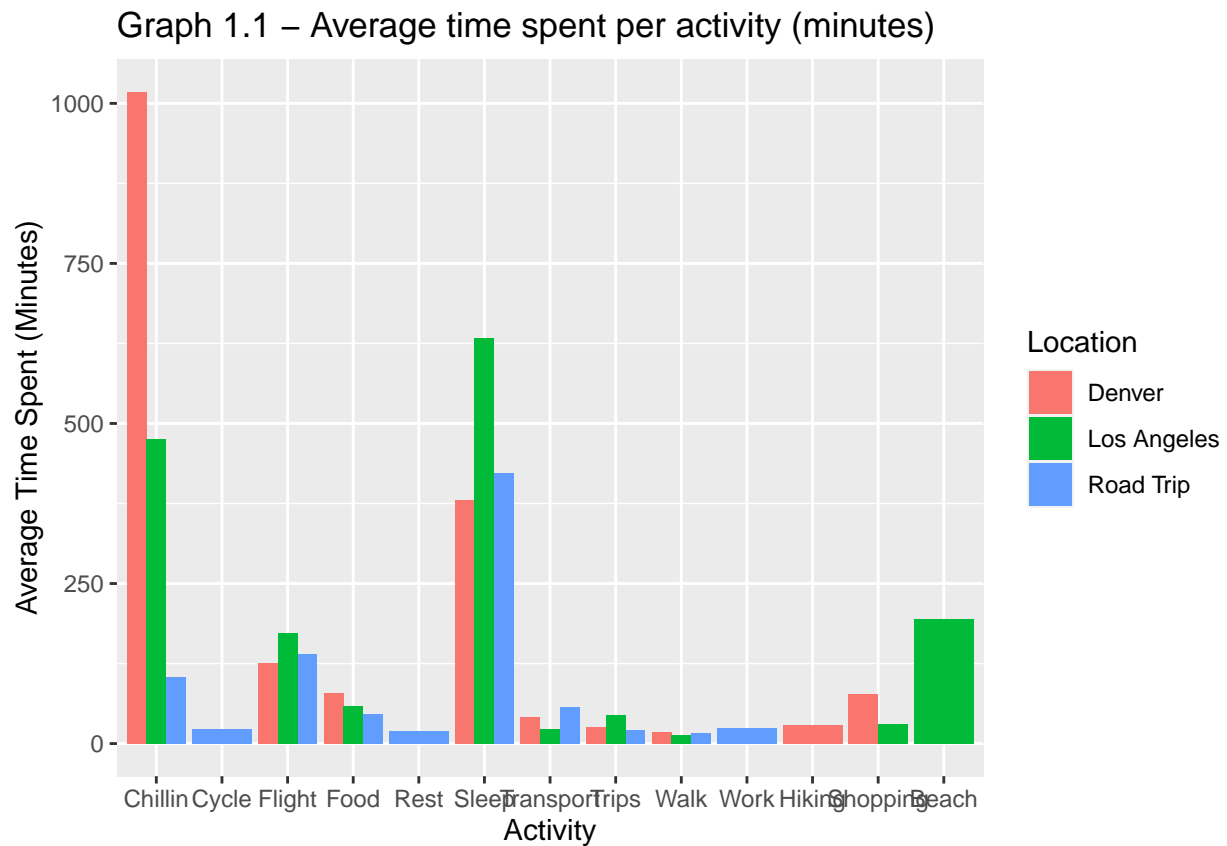

Graph 1.1 – Average time spent per activity (minutes)

```
print("This graph shows how much time I spent, on average, for each activity listed throughout the diff

## [1] "This graph shows how much time I spent, on average, for each activity listed throughout the dif
```

Describe what information is conveyed through data visualization #2 (that you'll create below) here.

```
library(ggplot2)

#Counting times each activity was documented by location
RT.tally = tally(group_by(logRT, NAME))
RT.tally$area = "Road Trip"
DEN.tally = tally(group_by(logDEN, NAME))
DEN.tally$area = "Denver"
LA.tally = tally(group_by(logLA, NAME))
LA.tally$area = "Los Angeles"

#Combining all 3 tables
tally = rbind(RT.tally, DEN.tally, LA.tally)

#Creating and printing graph
tally = ggplot(tally, aes(fill=area, x=NAME, y=n)) + geom_bar(stat="identity", position="dodge") +
  labs(x="Activity", y="Frequency of Activity", title = "Graph 2.1 - Times each  activity was documente
tally
```


Graph 2.1 – Times each activity was documented by location
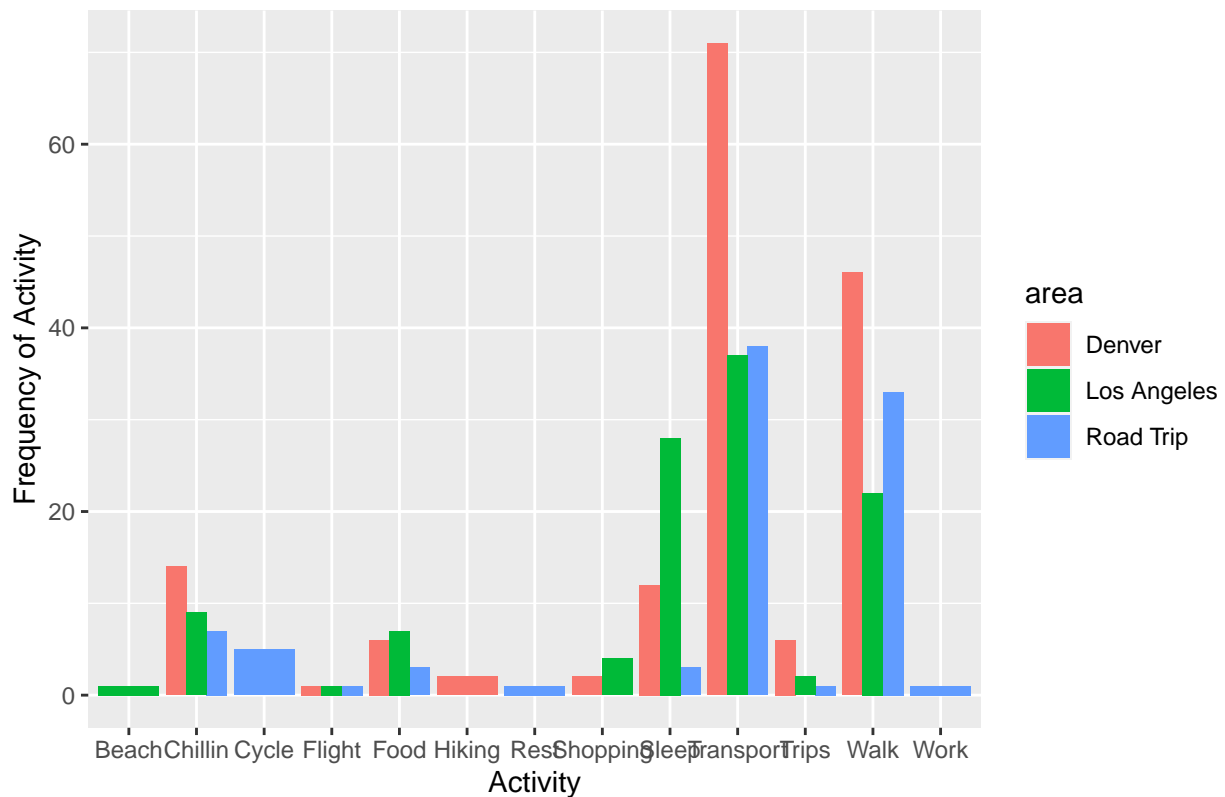
```r
print("This graph shows how many times I did each activity listed throughout the different areas of my
```

```
## [1] "This graph shows how many times I did each activity listed throughout the different areas of my
```

Describe what information is conveyed through the table (that you'll create below) here.

```r
#Selecting most important columns from larger data table
present = subset(log, select = -c(3:5, 8, 10))
present
```

A tibble: 364 x 5

START.DATE END.DATE MINUTES NAME general.location
1 2020-08-15 10:37:42 2020-08-15 13:28:30 171. Transport Road Trip
2 2020-08-15 13:28:30 2020-08-15 13:57:55 29.4 Chillin Road Trip
3 2020-08-15 13:57:55 2020-08-15 15:34:27 96.5 Transport Road Trip
4 2020-08-15 15:34:27 2020-08-15 15:39:48 5.35 Walk Road Trip
5 2020-08-15 15:47:25 2020-08-15 17:08:09 80.7 Transport Road Trip
6 2020-08-15 17:08:09 2020-08-15 17:27:07 19.0 Rest Road Trip
7 2020-08-15 17:27:07 2020-08-15 18:44:41 77.6 Transport Road Trip
8 2020-08-15 18:58:36 2020-08-15 20:13:30 74.9 Transport Road Trip
9 2020-08-15 20:13:30 2020-08-15 20:36:31 23.0 Food Road Trip
10 2020-08-15 20:36:31 2020-08-15 20:43:21 6.83 Walk Road Trip
... with 354 more rows

```r
print("This table shows each instance of my location chaning enough that the app noticed it, and documen
```

[1] "This table shows each instance of my location chaning enough that the app noticed it, and documented it as a different activity."

To conclude, briefly summarize what you found in response to the questions posed here.

Reflection

Write your one-page reflection here in paragraph form. In particular, address:

- What difficulties in the data collection and analysis process did you encounter? Identify two of your main hurdles in gathering accurate data. One problem was that the app from which I drew my data only collected data as I changed location, as a result I had to add in data when I didn't change location but changed activity. This created issues as it is more difficult to recall specific times when I changed what I was doing. I also found that this app gave me way too much data, including the exact seconds when I started and stopped activities and had data dating back to the summer of 2017.

- What implications does that have for future data collection and/or analysis projects? Essentially just that I should update my data more frequently and rely less on apps. I also wish I had used a service which was only intended for this project, as using an app which I've used since 2017 resulted in an abundance of data and titles of activities such as "Chillin" which makes sense to me but in academia is a rather inappropriate title.

- How much data do you think you'd need to collect in order to answer your question(s) of interest? Would it be hard to collect that data? Why or why not? I think the data I collected was sufficient, however it would have been nice to have been able to include more data. One set of data I had hoped to use at one point was my heart rate as measured by my watch. I discovered, however, that the readings were not consistent enough to look at alongside the day-to-day data I had analyzed.

- As someone who provides data, what expectations do you have when you give your data (e.g. to Facebook, Google, MapMyRun, etc.)?
Well I would hope that it stays private, but I think we all know that this is not the case. I think at the very least, these large corporations could be more transparent and cooperative about what data they share. There are some facets of my data that I don't mind being public, especially if I get my internet experienced tailored more towards my taste as a result, however I don't enjoy when google maps knows when and where my flights are and where I will be staying and the duration of my stays.

- As someone who analyzes others' data, what ethical responsibilities do you have? To analyze the data from an impartial view and, I believe, to use the data in an effort to better somebody's awareness of their own data. An important responsibility is to keep the data confidential and to communicate with the other party about your findings and any potential concerns regarding the data.