Assignment_Week_7&8_Venkidusamy_KesavAdithya

Kesav Adithya Venkidusamy

2022/07/22

```
knitr::opts_chunk$set(echo = TRUE)

library(readxl)
library(ggplot2)
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
```

Data Loading

```
# Creating dataframe
ppg_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 7&8/ppg2008.csv")
head(ppg_df)
```

```
##              Name  G  MIN  PTS  FGM  FGA   FGP FTM FTA   FTP X3PM X3PA  X3PP ORB
## 1     Dwyane Wade 79 38.6 30.2 10.8 22.0 0.491 7.5 9.8 0.765  1.1  3.5 0.317 1.1
## 2    LeBron James 81 37.7 28.4  9.7 19.9 0.489 7.3 9.4 0.780  1.6  4.7 0.344 1.3
## 3     Kobe Bryant 82 36.2 26.8  9.8 20.9 0.467 5.9 6.9 0.856  1.4  4.1 0.351 1.1
## 4   Dirk Nowitzki 81 37.7 25.9  9.6 20.0 0.479 6.0 6.7 0.890  0.8  2.1 0.359 1.1
## 5   Danny Granger 67 36.2 25.8  8.5 19.1 0.447 6.0 6.9 0.878  2.7  6.7 0.404 0.7
## 6    Kevin Durant 74 39.0 25.3  8.9 18.8 0.476 6.1 7.1 0.863  1.3  3.1 0.422 1.0
##   DRB TRB AST STL BLK  TO  PF
## 1 3.9 5.0 7.5 2.2 1.3 3.4 2.3
## 2 6.3 7.6 7.2 1.7 1.1 3.0 1.7
## 3 4.1 5.2 4.9 1.5 0.5 2.6 2.3
## 4 7.3 8.4 2.4 0.8 0.8 1.9 2.2
## 5 4.4 5.1 2.7 1.0 1.4 2.5 3.1
## 6 5.5 6.5 2.8 1.3 0.7 3.0 1.8
```

```
# Total number of records present in the data set
nrow(ppg_df)
```
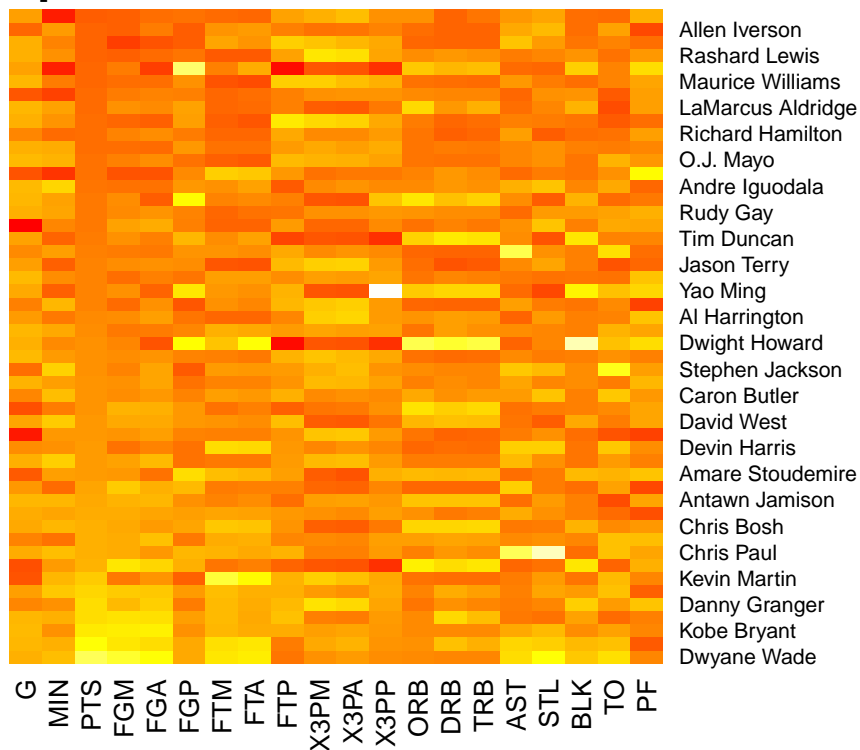
```
## [1] 50
```

```
# Scatter Plot
df1 <- data.frame(ppg_df[,-1], row.names = ppg_df[,1])
heatmap(as.matrix(df1), scale="column",col=heat.colors(100),main="              R: Heat Map Chart to show
```
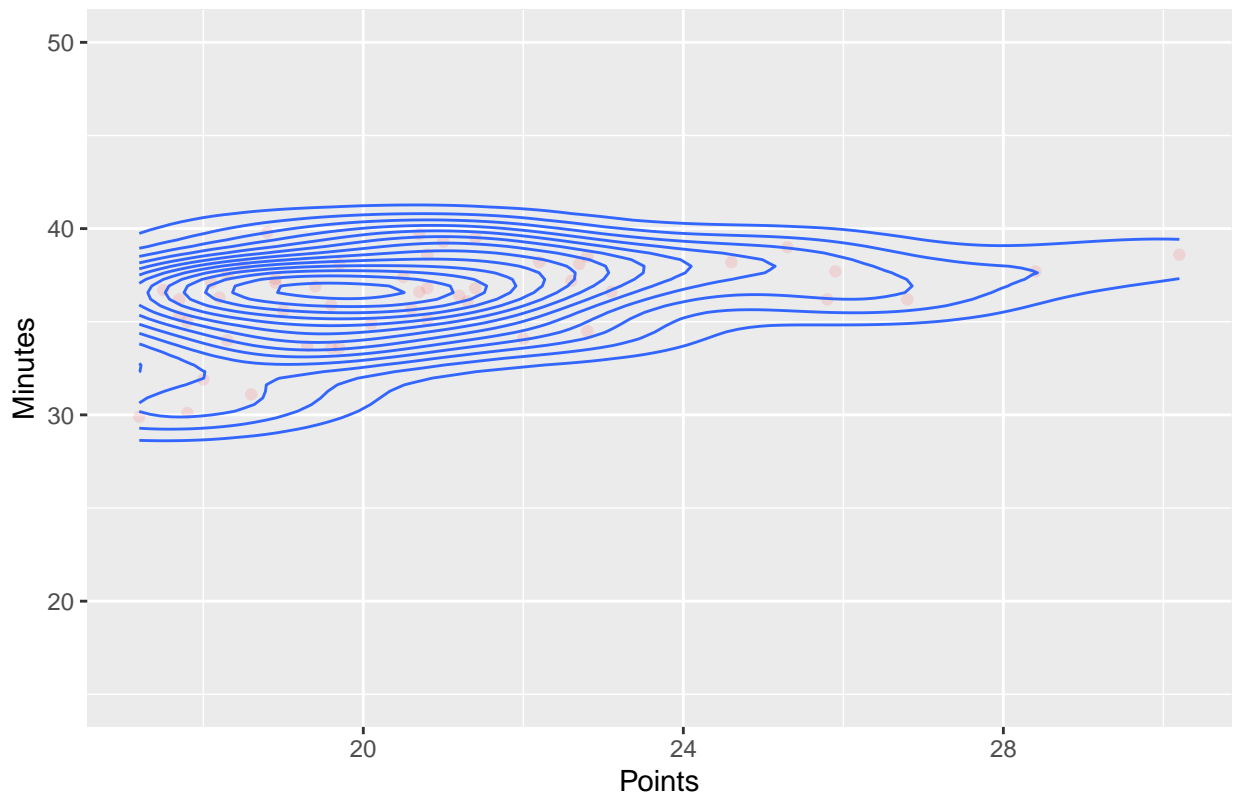
R: Heat Map Chart to show NBA Per Game Performance


```
## Create Contour Chart
```

```
ggplot(data=ppg_df, aes(x=PTS, y=MIN)) +
  ylim(15,50) +
  geom_point(alpha=0.1, col="red") +
  geom_density_2d() +
  ggtitle("R - Contour plot for PTS and MIN") +
  theme(plot.title = element_text(hjust=0.5)) +
  labs(x="Points", y="Minutes")
```

R – Contour plot for PTS and MIN


```r
# Creating dataframe
costco_df <- read.csv("E:/Personal/Bellevue University/Course/github/dsc640/Week 7&8/costcos-geocoded.cs
head(costco_df)
```

```
##                      Address        City    State   Zip.Code Latitude  Longitude
## 1  1205 N. Memorial Parkway  Huntsville  Alabama 35801-5930 34.74309   -86.60096
## 2        3650 Galleria Circle     Hoover  Alabama 35244-2346 33.37765   -86.81242
## 3     8251 Eastchase Parkway Montgomery  Alabama       36117 32.36389   -86.15088
## 4 5225 Commercial Boulevard      Juneau   Alaska 99801-7210 58.35920 -134.48300
## 5        330 West Dimond Blvd  Anchorage   Alaska 99515-1950 61.14327 -149.88422
## 6            4125 DeBarr Road  Anchorage   Alaska 99508-3115 61.21081 -149.80434
```

```r
# Total number of records present in the data set
nrow(costco_df)
```

```
## [1] 417
```

```r
library(maps)
```

```
## Warning: package 'maps' was built under R version 4.1.3
```

```r
library(mapdata)
```

```
## Warning: package 'mapdata' was built under R version 4.1.3
```

```
usa <- map_data("usa")

cost <- costco_df[costco_df$Longitude > -130,]
gg1 <- ggplot() +
  geom_polygon(data = usa, aes(x=long, y = lat, group = group), fill = "lightblue", color = "blue") +
  coord_fixed(1.3)
gg1 +
  geom_point(data=cost, aes(x=Longitude,y=Latitude), color="black",size=2)+
  geom_point(data=cost, aes(x=Longitude,y=Latitude), color="blue",size=1)+
  ggtitle("R - Saptial Map For Costco Store Locations")
```


R – Saptial Map For Costco Store Locations
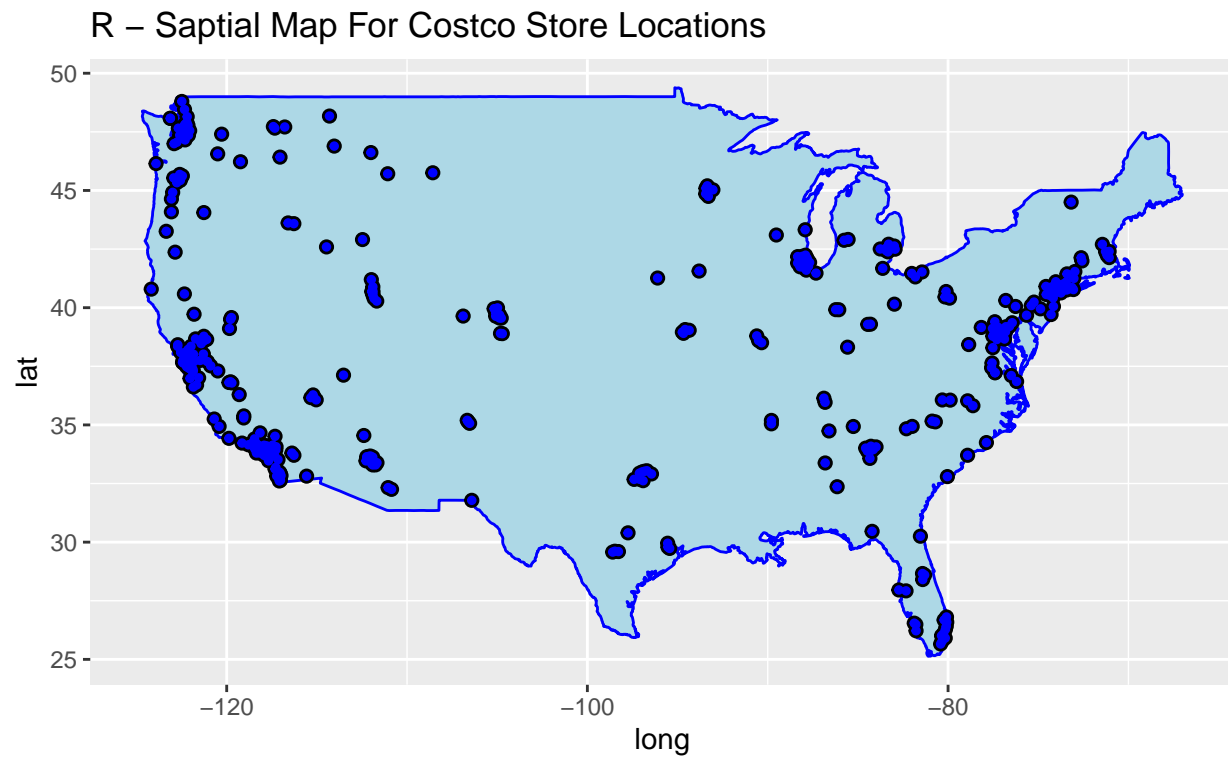
Week 7 & 8 Assignment - Python

Name: Kesav Adithya Venkidusamy
Course: DSC640 - Data Presentation and Visualization
Instructor: Catherine Williams

These two weeks we are going to be focused on heat maps, spatial charts, and contour charts and using various tools to create these visualizations. You must consolidate all the charts into ONE document with each chart labeled with the type of chart and technology - for example: Python - Bar Chart. Failure to label and consolidate the charts will resort in points being taken off or a 0 for the assignment.

Sample Datasets (click on the Downloads tab.)

You may also download them directly from this link: Exercise 4.2 Datasets (click the link to download a folder containing the datasets.)

You need to submit:
1 heat map, 1 spatial chart and 1 funnel or violin chart using Tableau or PowerBI
1 heat map, 1 spatial chart and 1 contour chart using Python
1 heat map, 1 spatial chart and 1 contour chart using R

1 heat map, 1 spatial chart and 1 contour chart using Python

In [57]:
```python
## Importing libraries required for this exercise
import pandas as pd
import numpy as np
import squarify
import matplotlib.pyplot as plt
%matplotlib inline
import plotly.express as px
import seaborn as sns
import matplotlib
import plotly.graph_objects as go
```

Read Input datasets

In [2]:
```python
## Reading the costco data
costco_df = pd.read_csv('costcos-geocoded.csv')
costco_df.head()
```

Out[2]:

	Address	City	State	Zip Code	Latitude	Longitude
0	1205 N. Memorial Parkway	Huntsville	Alabama	35801-5930	34.743095	-86.600955
1	3650 Galleria Circle	Hoover	Alabama	35244-2346	33.377649	-86.812420
2	8251 Eastchase Parkway	Montgomery	Alabama	36117	32.363889	-86.150884
3	5225 Commercial Boulevard	Juneau	Alaska	99801-7210	58.359200	-134.483000

	Address	City	State	Zip Code	Latitude	Longitude
4	330 West Dimond Blvd	Anchorage	Alaska	99515-1950	61.143266	-149.884217

In [22]:
```python
## Reading ppg dataset
ppg2008_df = pd.read_csv('ppg2008.csv')
ppg2008_df.head()
```

Out[22]:

	Name	G	MIN	PTS	FGM	FGA	FGP	FTM	FTA	FTP	...	3PA	3PP	ORB	DRB	TRB	AST
0	Dwyane Wade	79	38.6	30.2	10.8	22.0	0.491	7.5	9.8	0.765	...	3.5	0.317	1.1	3.9	5.0	7.5
1	LeBron James	81	37.7	28.4	9.7	19.9	0.489	7.3	9.4	0.780	...	4.7	0.344	1.3	6.3	7.6	7.2
2	Kobe Bryant	82	36.2	26.8	9.8	20.9	0.467	5.9	6.9	0.856	...	4.1	0.351	1.1	4.1	5.2	4.9
3	Dirk Nowitzki	81	37.7	25.9	9.6	20.0	0.479	6.0	6.7	0.890	...	2.1	0.359	1.1	7.3	8.4	2.4
4	Danny Granger	67	36.2	25.8	8.5	19.1	0.447	6.0	6.9	0.878	...	6.7	0.404	0.7	4.4	5.1	2.7

5 rows × 21 columns

In [23]:
```python
ppg2008_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 21 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Name    50 non-null     object
 1   G       50 non-null     int64
 2   MIN     50 non-null     float64
 3   PTS     50 non-null     float64
 4   FGM     50 non-null     float64
 5   FGA     50 non-null     float64
 6   FGP     50 non-null     float64
 7   FTM     50 non-null     float64
 8   FTA     50 non-null     float64
 9   FTP     50 non-null     float64
 10  3PM     50 non-null     float64
 11  3PA     50 non-null     float64
 12  3PP     50 non-null     float64
 13  ORB     50 non-null     float64
 14  DRB     50 non-null     float64
 15  TRB     50 non-null     float64
 16  AST     50 non-null     float64
 17  STL     50 non-null     float64
 18  BLK     50 non-null     float64
 19  TO      50 non-null     float64
 20  PF      50 non-null     float64
dtypes: float64(19), int64(1), object(1)
memory usage: 8.3+ KB
```

1. Python - Heat Map

Plotting Heat Map for ppg test result

In [34]:

```python
ppg2008_sub_df = ppg2008_df.copy()
ppg2008_sub_df.set_index('Name', inplace = True)
ppg2008_sub_df.head()
```

Out[34]:

Name	G	MIN	PTS	FGM	FGA	FGP	FTM	FTA	FTP	3PM	3PA	3PP	ORB	DRB	TRB	AST
Dwyane Wade	79	38.6	30.2	10.8	22.0	0.491	7.5	9.8	0.765	1.1	3.5	0.317	1.1	3.9	5.0	7.
LeBron James	81	37.7	28.4	9.7	19.9	0.489	7.3	9.4	0.780	1.6	4.7	0.344	1.3	6.3	7.6	7.
Kobe Bryant	82	36.2	26.8	9.8	20.9	0.467	5.9	6.9	0.856	1.4	4.1	0.351	1.1	4.1	5.2	4.
Dirk Nowitzki	81	37.7	25.9	9.6	20.0	0.479	6.0	6.7	0.890	0.8	2.1	0.359	1.1	7.3	8.4	2.
Danny Granger	67	36.2	25.8	8.5	19.1	0.447	6.0	6.9	0.878	2.7	6.7	0.404	0.7	4.4	5.1	2.

In [39]:

```python
# Normalize columns
ppg2008_sub_norm = (ppg2008_sub_df - ppg2008_sub_df.mean())/ppg2008_sub_df.std()
```

In [51]:

```python
# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(13, 12))

# Create heatmap of normalized data
sns.heatmap(ppg2008_sub_norm, cmap='YlOrRd', cbar=False)

# Add chart title and labels
plt.title("Python - Heat Map to show NBA Per Game Performance", fontsize = 15, loc = 'l
plt.ylabel("")

plt.show()
```

Python - Heat Map to show NBA Per Game Performance
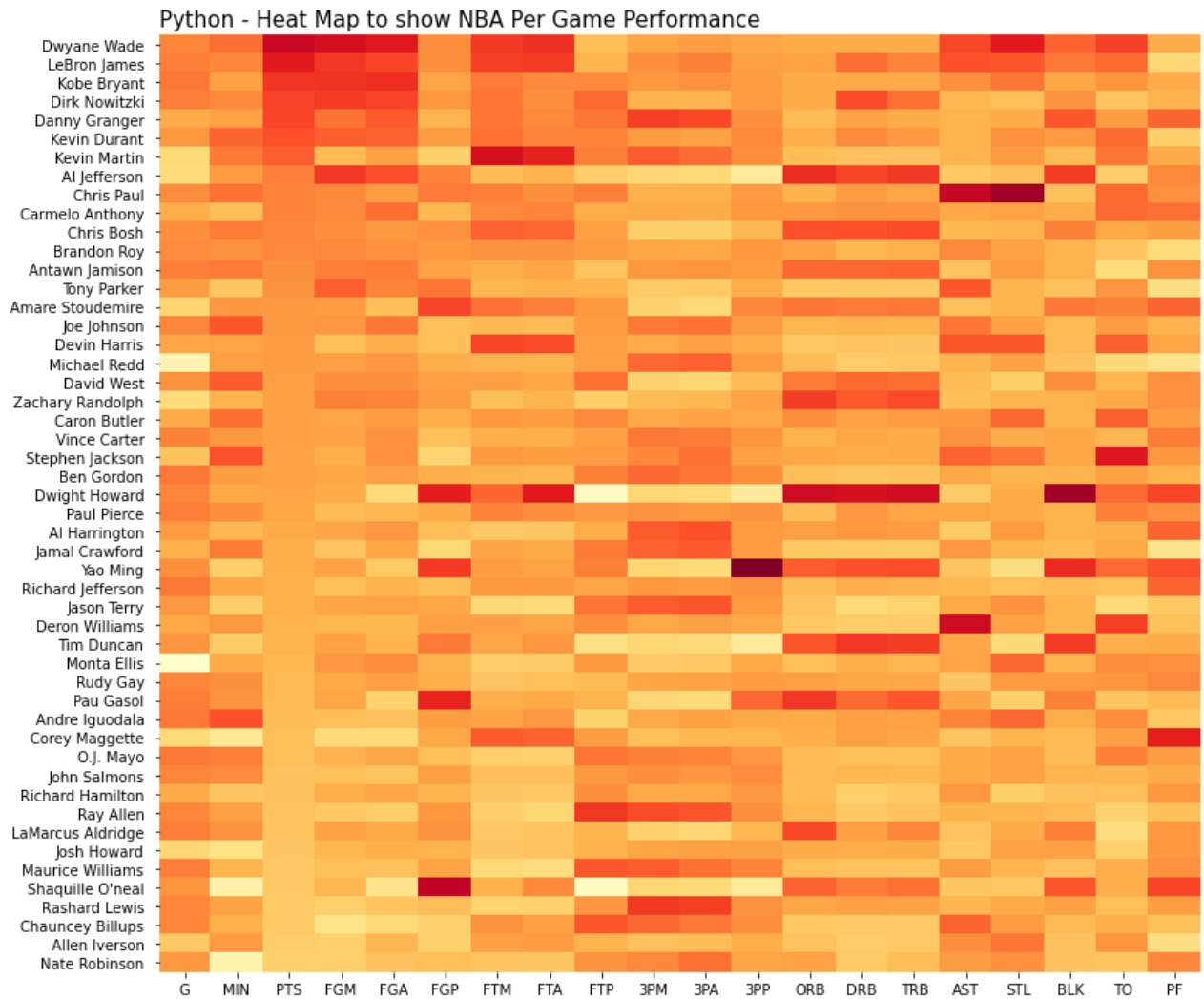
Python - Spatial Chart

```
In [65]:   fig = go.Figure(data=go.Scattergeo(
                   locationmode = 'USA-states',
                   lon = costco_df['Longitude'],
                   lat = costco_df['Latitude'],
                   text = costco_df['Address'],
                   mode = 'markers',
                   marker = dict(
                       size = 8,
                       opacity = 0.8,
                       reversescale = True,
                       autocolorscale = False,
                       symbol = 'circle',
                       line = dict(
                               width = 1,
                               color = 'rgba(102, 102, 102)'
                               ),
                       colorscale = 'Blues',
                       cmin = 0,
                       colorbar_title = "Costco Store Locations"
                       )))

           fig.update_layout(
                   title = 'Python - Spatial Chart to show Costco Locations',
```

```
                geo = dict(
                        scope = 'usa',
                        projection_type = 'albers usa',
                        showland = True,
                        landcolor = "rgb(250, 250, 250)",
                        subunitcolor = "rgb(217, 217, 217)",
                        countrycolor = "rgb(217, 217, 217)",
                        countrywidth = 0.5,
                        subunitwidth = 0.5
                        ),
                )

        fig.show()
```

Python - Spatial Chart to show Costco Locations
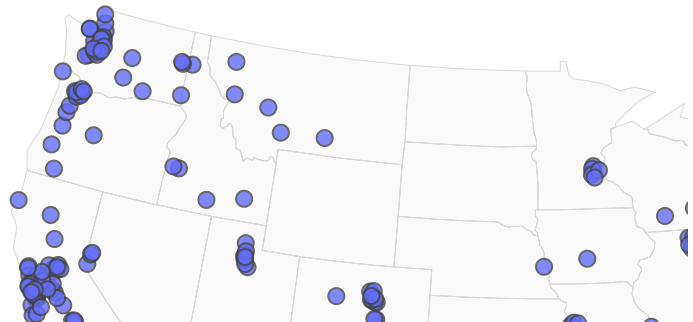
Python - Contour Chart

```
In [66]:  ## Initialize the matplotlib figure
          f, ax = plt.subplots(figsize=(10,6))

          sns.kdeplot(x=costco_df.Longitude, y = costco_df.Latitude,
                      cmap = "Blues", shade = True, thresh=0)

          ## Add chart title and labels
          plt.title("Python - Contour Plot of Costco Location by Latitude and Logitude",
```

```
                fontsize = 15,loc = 'left')
plt.xlabel('Logtitude',fontsize =12)
plt.ylabel('Latitude', fontsize = 12)

plt.show()
```


Python - Contour Plot of Costco Location by Latitude and Logitude

In [ ]:

Tableau: Heat Map to Show NBA Game Performance

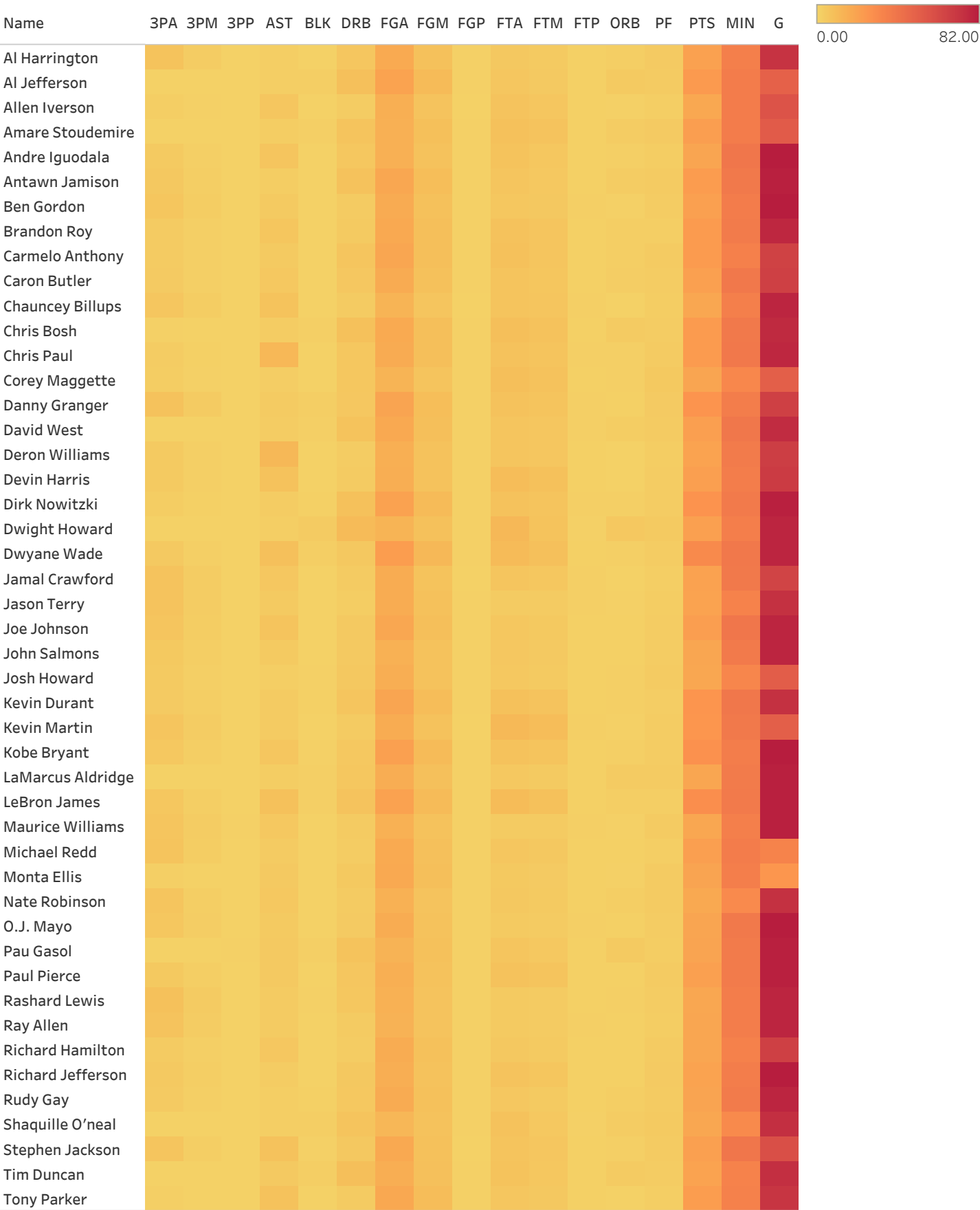| Name	3PA	3PM	3PP	AST	BLK	DRB	FGA	FGM	FGP	FTA	FTM	FTP	ORB	PF	PTS	MIN	G
Al Harrington																	
Al Jefferson																	
Allen Iverson																	
Amare Stoudemire																	
Andre Iguodala																	
Antawn Jamison																	
Ben Gordon																	
Brandon Roy																	
Carmelo Anthony																	
Caron Butler																	
Chauncey Billups																	
Chris Bosh																	
Chris Paul																	
Corey Maggette																	
Danny Granger																	
David West																	
Deron Williams																	
Devin Harris																	
Dirk Nowitzki																	
Dwight Howard																	
Dwyane Wade																	
Jamal Crawford																	
Jason Terry																	
Joe Johnson																	
John Salmons																	
Josh Howard																	
Kevin Durant																	
Kevin Martin																	
Kobe Bryant																	
LaMarcus Aldridge																	
LeBron James																	
Maurice Williams																	
Michael Redd																	
Monta Ellis																	
Nate Robinson																	
O.J. Mayo																	
Pau Gasol																	
Paul Pierce																	
Rashard Lewis																	
Ray Allen																	
Richard Hamilton																	
Richard Jefferson																	
Rudy Gay																	
Shaquille O'neal																	
Stephen Jackson																	
Tim Duncan																	
Tony Parker																	

Color scale: 0.00 — 82.00

Sum of Pivot Field Values (color) broken down by Pivot Field Names vs. Name. The view is filtered on Pivot Field Names, which excludes STL, TO and TRB.

Tableau: Heat Map to Show NBA Game Performance
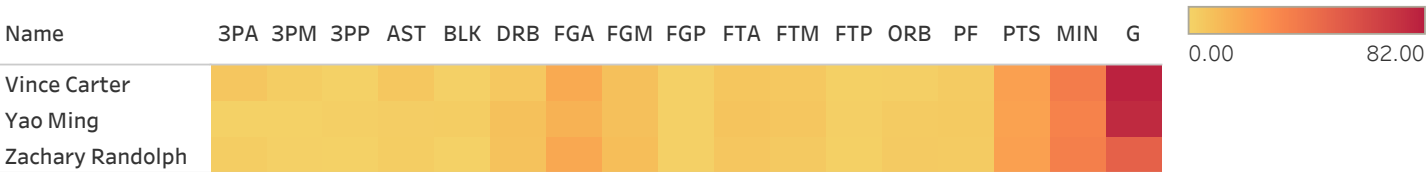
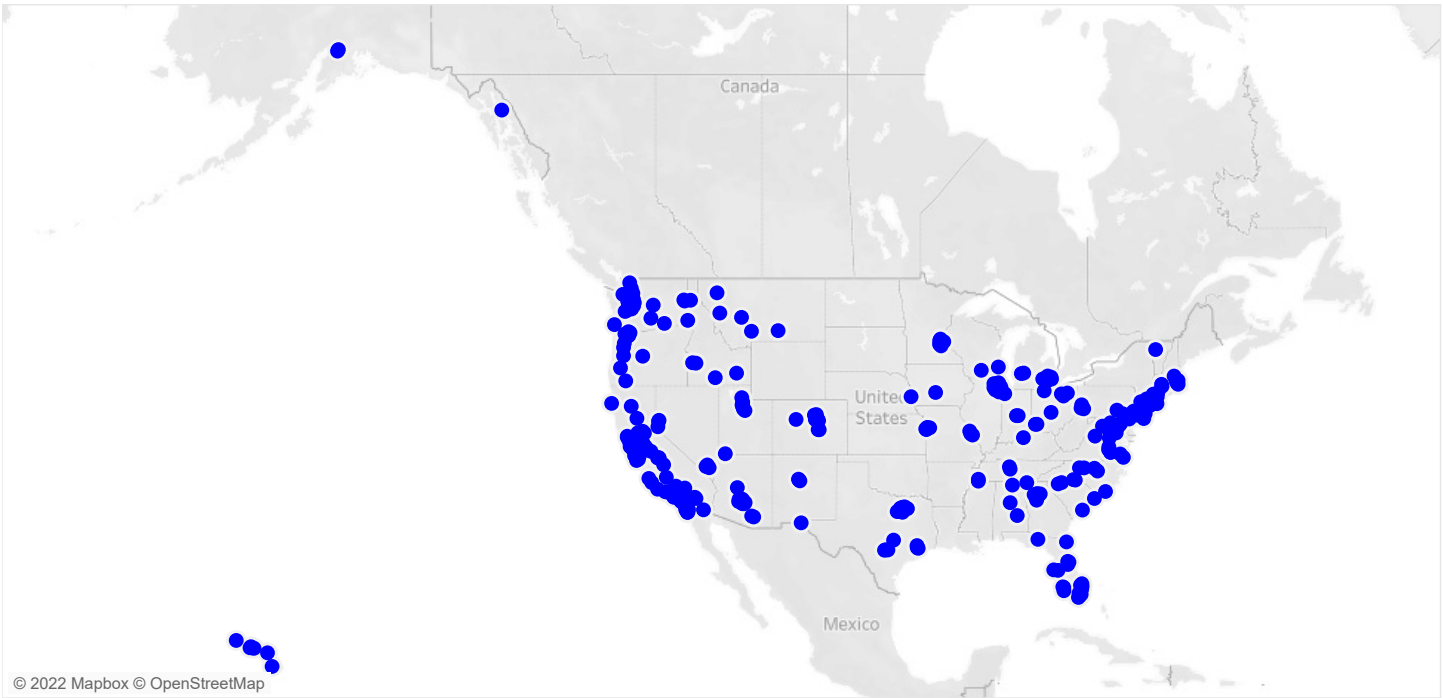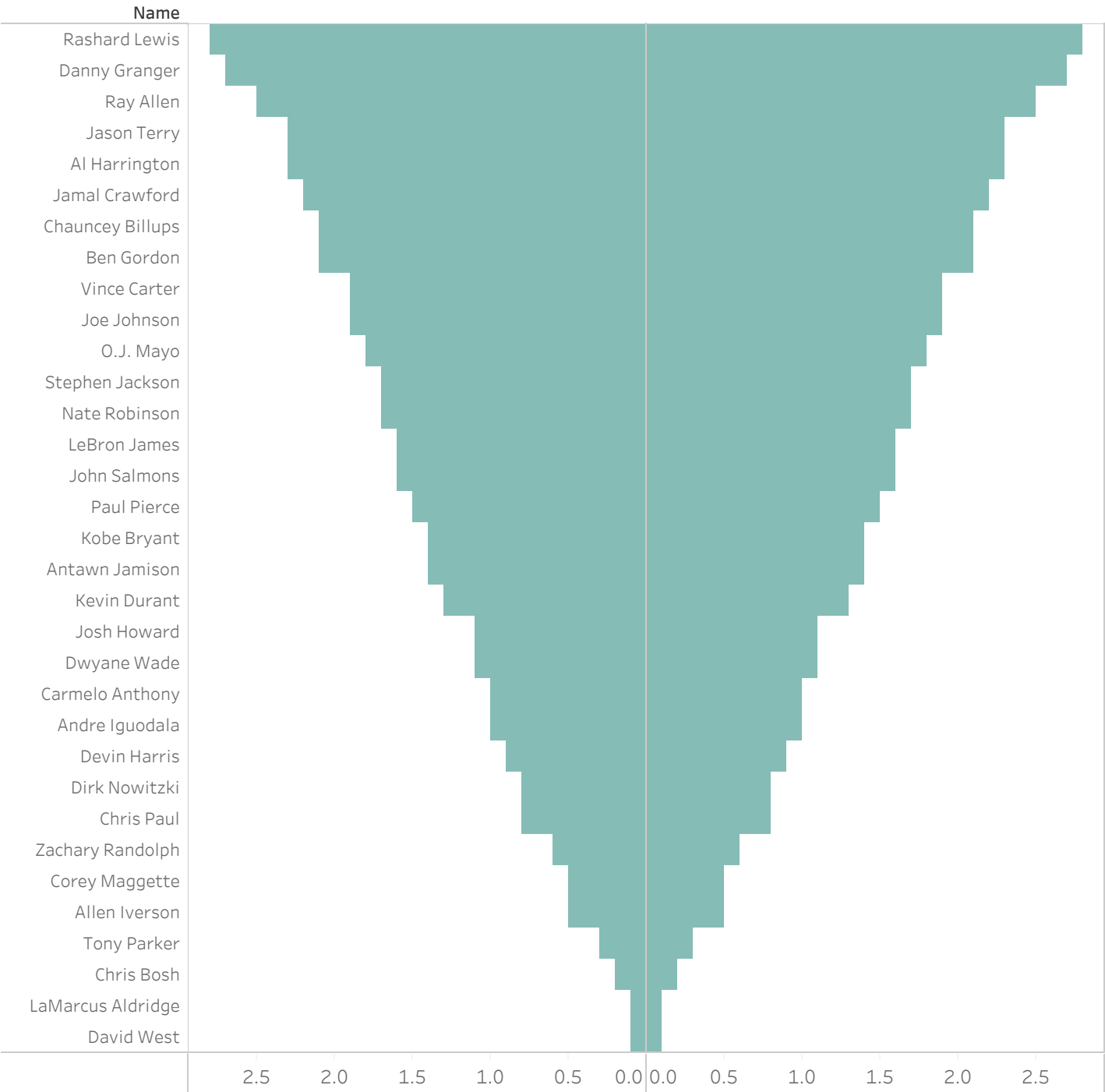| Name	3PA	3PM	3PP	AST	BLK	DRB	FGA	FGM	FGP	FTA	FTM	FTP	ORB	PF	PTS	MIN	G
Vince Carter																	
Yao Ming																	
Zachary Randolph																	

0.00   82.00

Sum of Pivot Field Values (color) broken down by Pivot Field Names vs. Name. The view is filtered on Pivot Field Names, which excludes STL, TO and TRB.

Tableau - Spatial Chart to show Costco Location across USA

Map based on average of Longitude and average of Latitude. Details are shown for Zip Code.

© 2022 Mapbox © OpenStreetMap

Tableau: Funnel Chart to show NBA for 3PM

Name		
Rashard Lewis		
Danny Granger		
Ray Allen		
Jason Terry		
Al Harrington		
Jamal Crawford		
Chauncey Billups		
Ben Gordon		
Vince Carter		
Joe Johnson		
O.J. Mayo		
Stephen Jackson		
Nate Robinson		
LeBron James		
John Salmons		
Paul Pierce		
Kobe Bryant		
Antawn Jamison		
Kevin Durant		
Josh Howard		
Dwyane Wade		
Carmelo Anthony		
Andre Iguodala		
Devin Harris		
Dirk Nowitzki		
Chris Paul		
Zachary Randolph		
Corey Maggette		
Allen Iverson		
Tony Parker		
Chris Bosh		
LaMarcus Aldridge		
David West		

2.5   2.0   1.5   1.0   0.5   0.00.0   0.5   1.0   1.5   2.0   2.5

Sum of 3PM and sum of 3PM for each Name. The view is filtered on sum of 3PM and Name. The sum of 3PM filter ranges from 0.100 to 2.800. The Name filter keeps 33 of 50 members.