Ryan Long
DSC640
5.2 Exercises: Charts

Dataset used:
costcos-geocoded.csv
ppg2008.csv

Summary
I think the heat maps were the least challenging to create out of this series of charts. The Power BI method did take a little bit longer but resulted in better contrast in the values in each of the columns. The Python and R methods don't offer much distinction between the values. Perhaps using less would offer greater variation and more functionality.

The Funnel Chart in Power BI did provide me some interpretation challenges. It was until after hovering over the categories that I saw the bar at the top indicating the top bar represented 100% and the bars below were all a % of California's value count of stores. In the future I would try to setup as absolute values instead of a % for quicker interpretation.

I spent the most time on the initial spatial chart. I had many troubles getting a package to load / work and once I did, I had to figure out how to frame up the map on the USA to appropriately show the datapoints. Once I had it figured out in Python it was a bit easier to perform in R.

The following pages contain:
Power BI:
Heat Map
Spatial Chart
Funnel Chart

Python:
Heat Map
Spatial Chart
Contour Chart

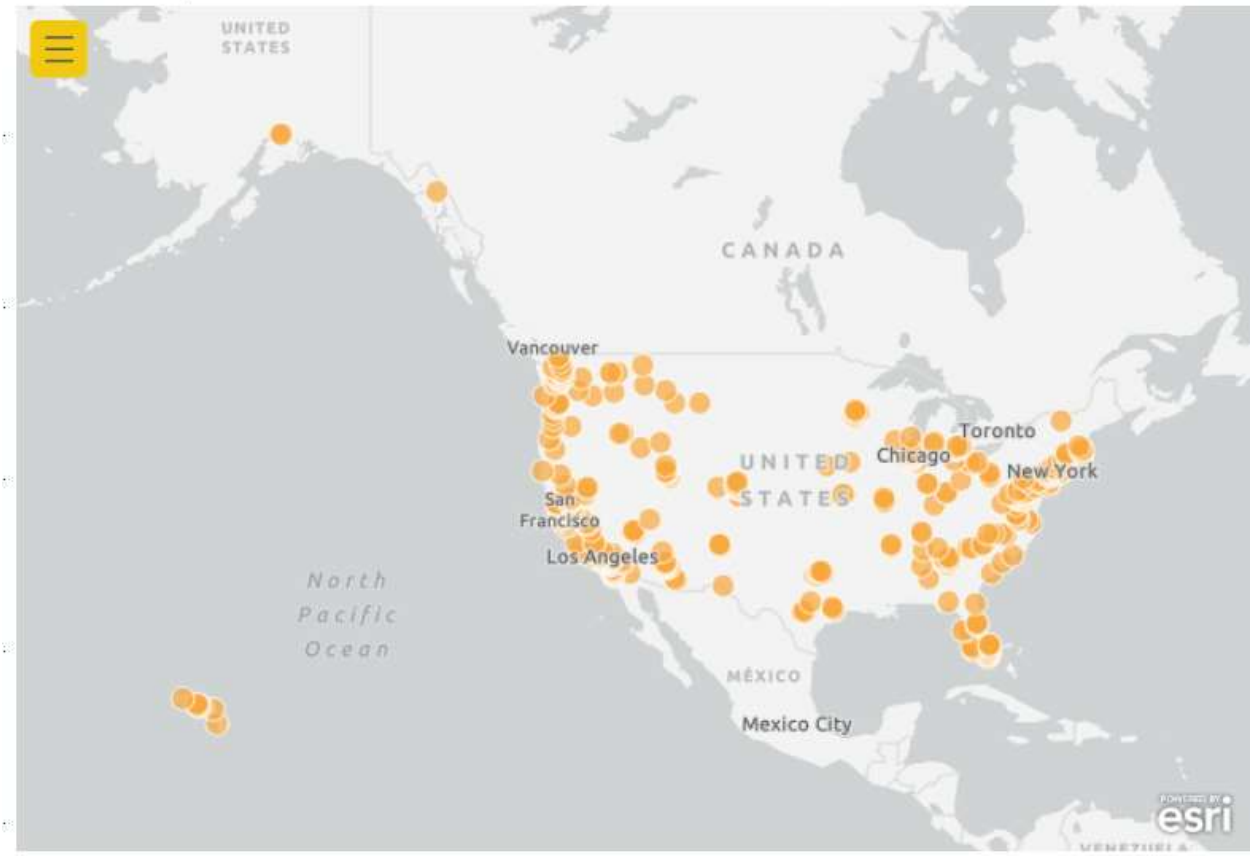
R:
Heat Map
Spatial Chart
Contour Chart

Appendix
Code support for both Python and R notebooks

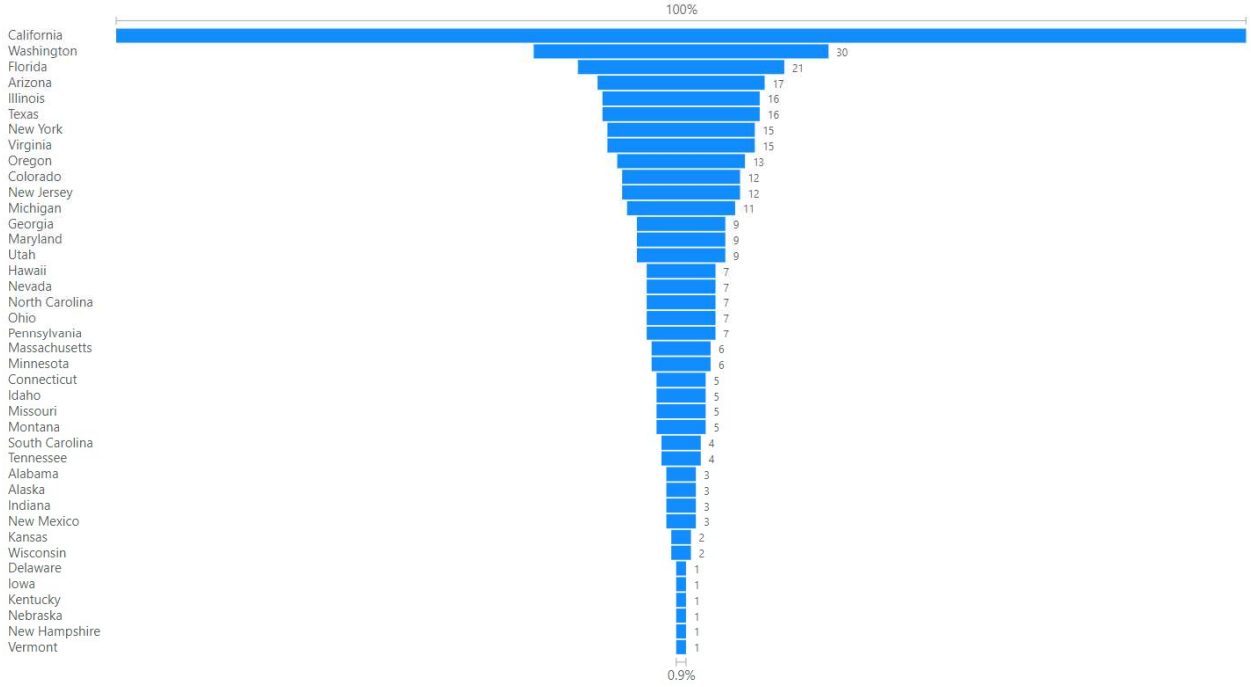
Power BI - Heat Map: NBA Player Stats 2008

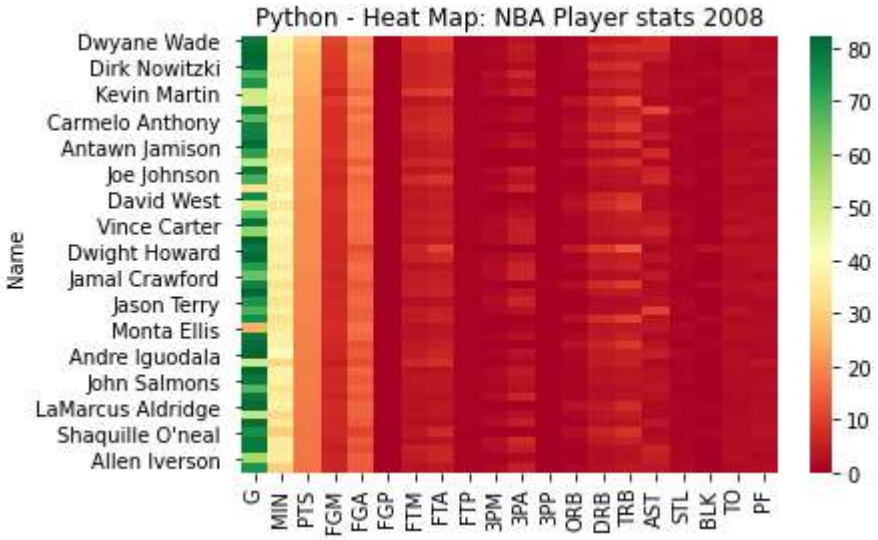
Name	G	MIN	PTS	FGM	FGA	FGP	FTM	FTA	FTP	3PM	3PA	3PP	ORB	DRB	TRB	AST	STL	TO	PF
Al Harrington	73	34.90	20.10	7.30	16.60	0.44	3.20	4.00	0.79	2.30	6.40	0.36	1.40	4.90	6.20	1.40	1.20	2.20	3.10
Al Jefferson	50	36.60	23.10	9.70	19.50	0.50	3.70	5.00	0.74	0.00	0.10	0.00	3.40	7.50	11.00	1.60	0.80	1.80	2.80
Allen Iverson	57	36.70	17.50	6.10	14.60	0.42	4.80	6.10	0.78	0.50	1.70	0.28	0.50	2.50	3.00	5.00	1.50	2.60	1.50
Amare Stoudemire	53	36.80	21.40	7.60	14.10	0.54	6.10	7.30	0.84	0.10	0.10	0.43	2.20	5.90	8.10	2.00	0.90	2.80	3.10
Andre Iguodala	82	39.80	18.80	6.60	14.00	0.47	4.60	6.40	0.72	1.00	3.20	0.31	1.10	4.60	5.70	5.30	1.60	2.70	1.90
Antawn Jamison	81	38.20	22.20	8.30	17.80	0.47	4.20	5.60	0.75	1.40	3.90	0.35	2.40	6.50	8.90	1.90	1.20	1.50	2.70
Ben Gordon	82	36.60	20.70	7.30	16.00	0.46	4.00	4.70	0.86	2.10	5.10	0.41	0.60	2.80	3.50	3.40	0.90	2.40	2.20
Brandon Roy	78	37.20	22.60	8.10	16.90	0.48	5.30	6.50	0.82	1.10	2.80	0.38	1.30	3.40	4.70	5.10	1.10	1.90	1.60
Carmelo Anthony	66	34.50	22.80	8.10	18.30	0.44	5.60	7.10	0.79	1.00	2.60	0.37	1.60	5.20	6.80	3.40	1.10	3.00	3.00
Caron Butler	67	38.60	20.80	7.30	16.20	0.45	5.10	6.00	0.86	1.00	3.10	0.31	1.80	4.40	6.20	4.30	1.60	3.10	2.50
Chauncey Billups	79	35.30	17.70	5.20	12.40	0.42	5.30	5.80	0.91	2.10	5.00	0.41	0.40	2.60	3.00	6.40	1.20	2.20	2.00
Chris Bosh	77	38.10	22.70	8.00	16.40	0.49	6.50	8.00	0.82	0.20	0.60	0.25	2.80	7.20	10.00	2.50	0.90	2.30	2.50
Chris Paul	78	38.50	22.80	8.10	16.10	0.50	5.80	6.70	0.87	0.80	2.30	0.36	0.90	4.70	5.50	11.00	2.80	3.00	2.70
Corey Maggette	51	31.10	18.60	5.70	12.40	0.46	6.70	8.10	0.82	0.50	1.90	0.25	1.00	4.60	5.50	1.80	0.90	2.40	3.80
Danny Granger	67	36.20	25.80	8.50	19.10	0.45	6.00	6.90	0.88	2.70	6.70	0.40	0.70	4.40	5.10	2.70	1.00	2.50	3.10
David West	76	39.30	21.00	8.00	17.00	0.47	4.80	5.50	0.88	0.10	0.30	0.24	2.10	6.40	8.50	2.30	0.60	2.10	2.70
Deron Williams	68	36.90	19.40	6.80	14.50	0.47	4.80	5.60	0.85	1.00	3.30	0.31	0.40	2.50	2.90	10.70	1.10	3.40	2.00

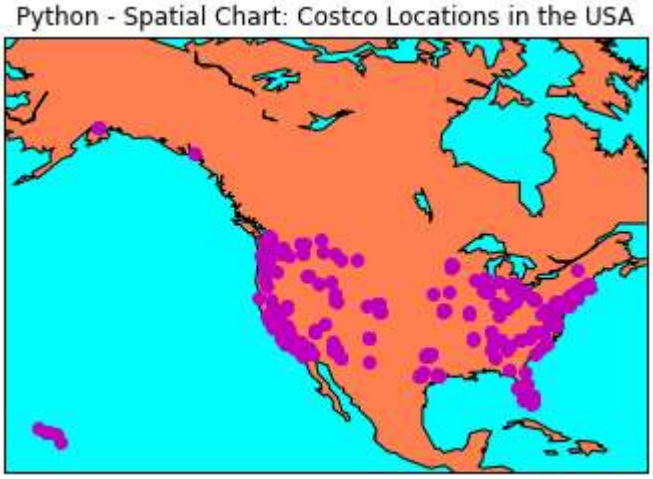
Power BI - Spatial Chart: Costco Locations in USA

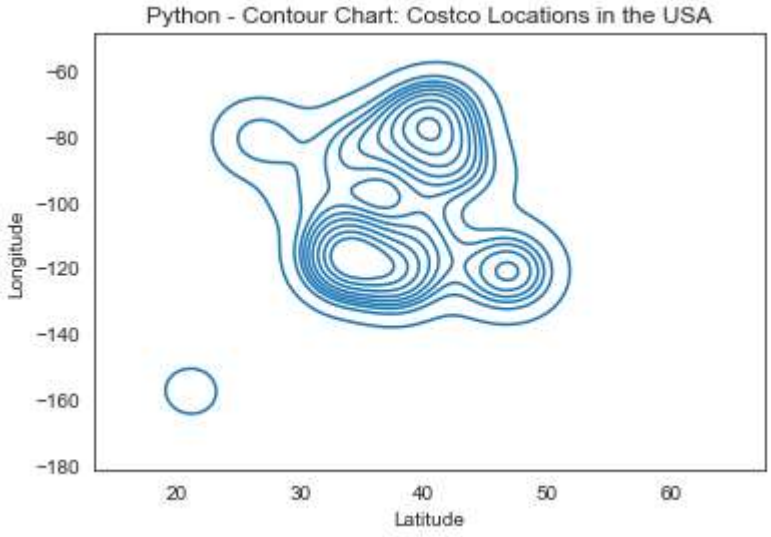


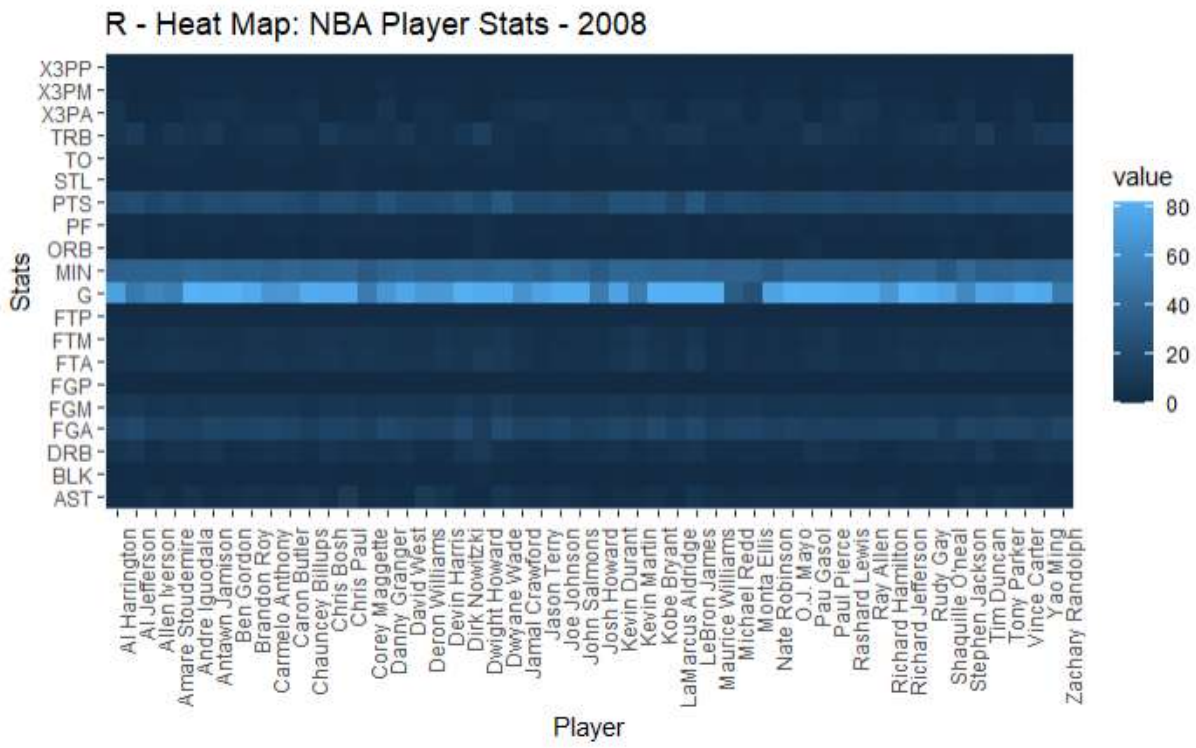
Power BI - Funnel Chart: Costco Locations by USA State

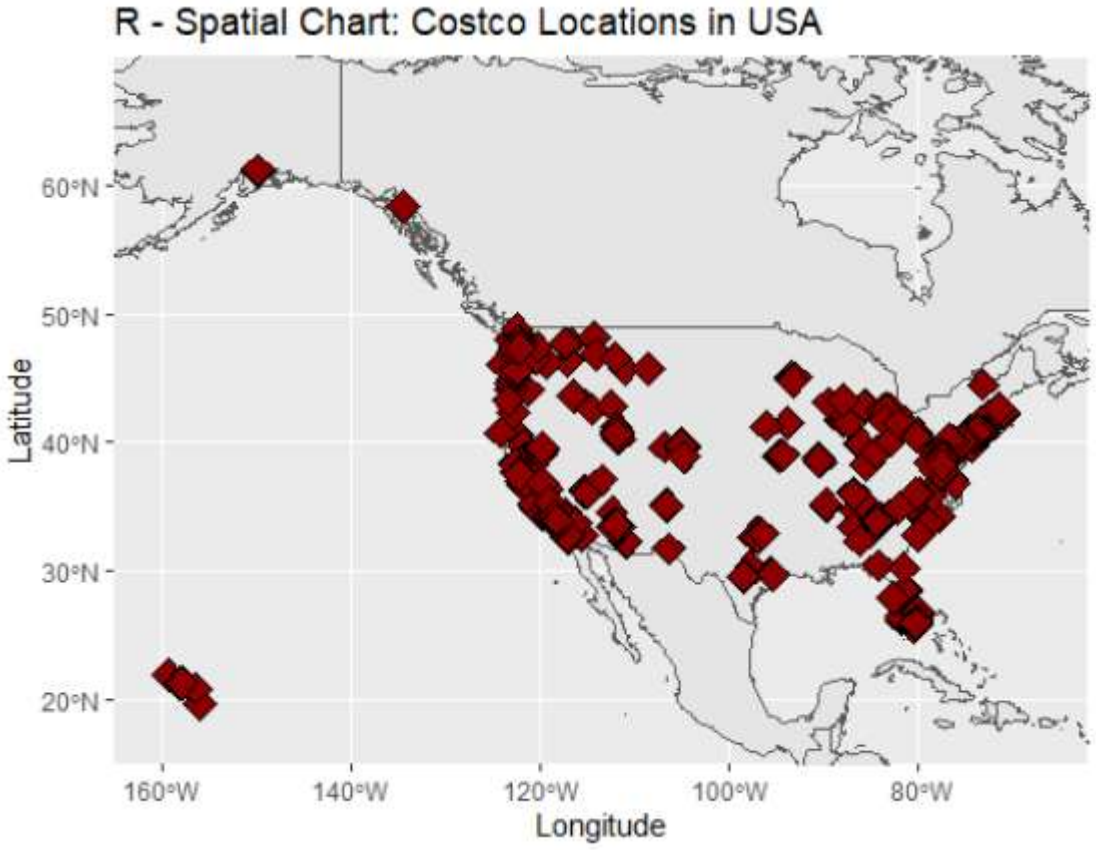


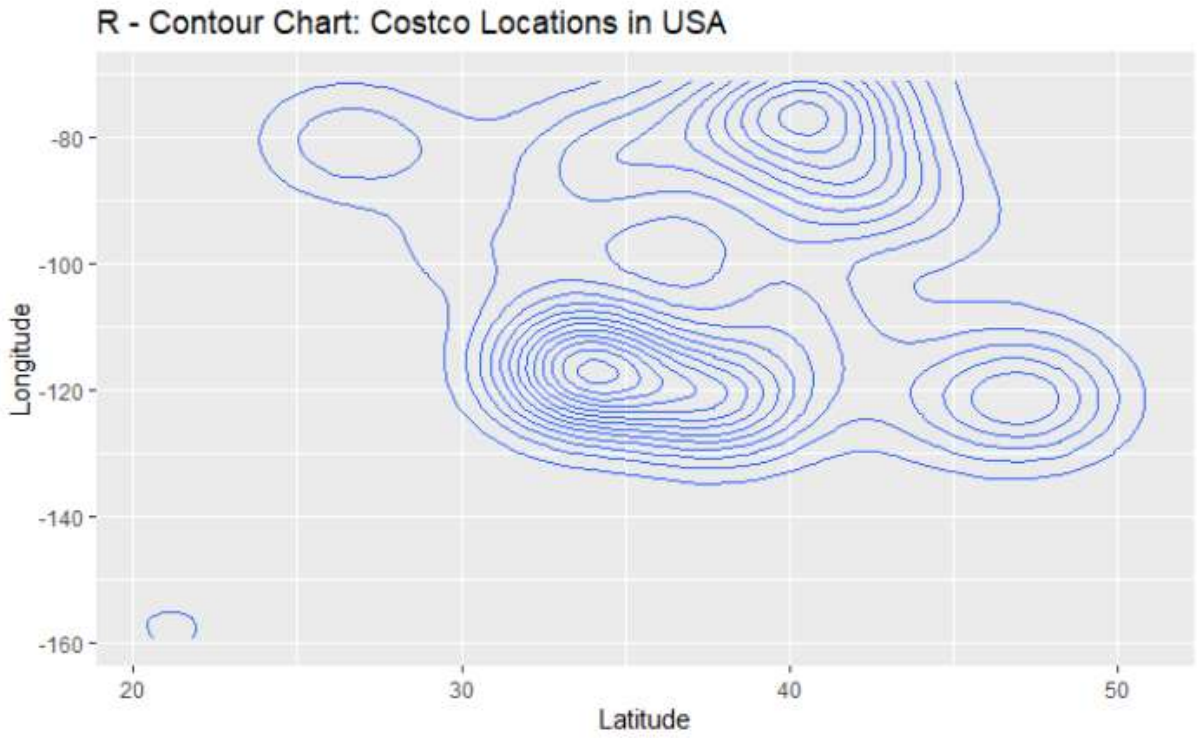

Python - Heat Map: NBA Player stats 2008
Dwyane Wade
Dirk Nowitzki
Kevin Martin
Carmelo Anthony
Antawn Jamison
Joe Johnson
David West
Vince Carter
Dwight Howard
Jamal Crawford
Jason Terry
Monta Ellis
Andre Iguodala
John Salmons
LaMarcus Aldridge
Shaquille O'neal
Allen Iverson
Name
G
MIN
PTS
FGM
FGA
FGP
FTM
FTA
FTP
3PM
3PA
3PP
ORB
DRB
TRB
AST
STL
BLK
TO
PF
80
70
60
50
40
30
20
10
0


Python - Spatial Chart: Costco Locations in the USA


Python - Contour Chart: Costco Locations in the USA
Longitude
−60
−80
−100
−120
−140
−160
−180
20
30
40
50
60
Latitude


R - Heat Map: NBA Player Stats - 2008
Stats
X3PP
X3PM
X3PA
TRB
TO
STL
PTS
PF
ORB
MIN
G
FTP
FTM
FTA
FGP
FGM
FGA
DRB
BLK
AST
value
80
60
40
20
0
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 Oneal
Stephen Jackson
Tim Duncan
Tony Parker
Vince Carter
Yao Ming
Zachary Randolph
Player


R - Spatial Chart: Costco Locations in USA
Latitude
60°N
50°N
40°N
30°N
20°N
160°W
140°W
120°W
100°W
80°W
Longitude


R - Contour Chart: Costco Locations in USA
-80
-100
-120
-140
-160
Longitude
20
30
40
50
Latitude

APPENDIX

In [1]:
```
#load libraries
import pandas as pd
import seaborn as sns

import numpy as np
import matplotlib.pyplot as plt

#import squarify
#import matplotlib.ticker as plticker # for plot ticks
```

In [2]:
```
#import data as dataframe

nba = pd.read_csv('ppg2008.csv', index_col=0)
costco = pd.read_csv('costcos-geocoded.csv') #tab seperator
```

In [3]:
```
nba.head()
```

Out[3]:

Name	G	MIN	PTS	FGM	FGA	FGP	FTM	FTA	FTP	3PM	3PA	3PP	ORB	DRB	TRB
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
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
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
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
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

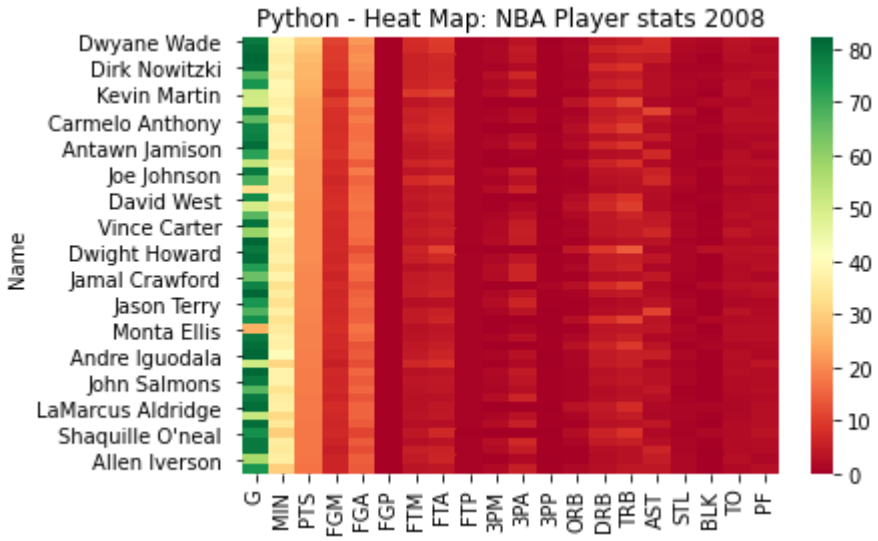
In [4]:
```
nba.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 50 entries, Dwyane Wade  to Nate Robinson
Data columns (total 20 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   G       50 non-null     int64  
 1   MIN     50 non-null     float64
 2   PTS     50 non-null     float64
 3   FGM     50 non-null     float64
 4   FGA     50 non-null     float64
 5   FGP     50 non-null     float64
 6   FTM     50 non-null     float64
 7   FTA     50 non-null     float64
 8   FTP     50 non-null     float64
 9   3PM     50 non-null     float64
 10  3PA     50 non-null     float64
 11  3PP     50 non-null     float64
 12  ORB     50 non-null     float64
 13  DRB     50 non-null     float64
 14  TRB     50 non-null     float64
 15  AST     50 non-null     float64
 16  STL     50 non-null     float64
 17  BLK     50 non-null     float64
 18  TO      50 non-null     float64
 19  PF      50 non-null     float64
dtypes: float64(19), int64(1)
memory usage: 8.2+ KB
```

Heatmap

In [5]:
```
# Default heatmap
sns.heatmap(nba,cmap ='RdYlGn')
plt.title("Python - Heat Map: NBA Player stats 2008")
```

Out[5]:
```
Text(0.5, 1.0, 'Python - Heat Map: NBA Player stats 2008')
```



Spatial Chart

In [6]:
```
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import numpy as np
```

In [7]:
```
costco
```

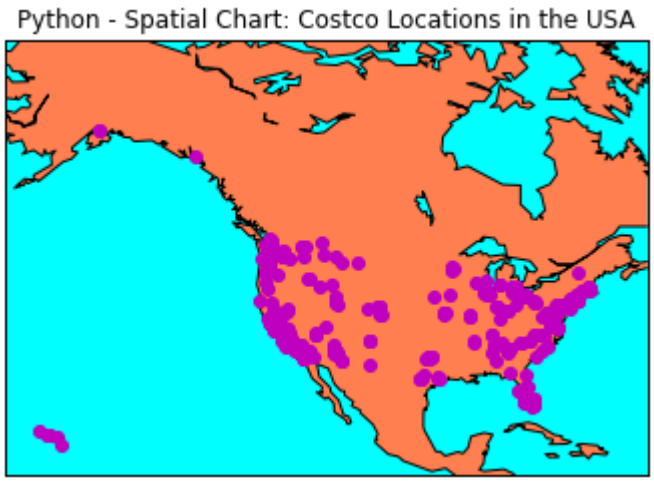
Out[7]:

	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
4	330 West Dimond Blvd	Anchorage	Alaska	99515-1950	61.143266	-149.884217
...	...	...	...	...	...	...
412	19610 SE 1st St	Vancouver	Washington	98607	45.621299	-122.459135
413	10990 Harbor Hill Dr	Gig Harbor	Washington	98335	47.357748	-122.603888
414	27520 Covington Way SE	Covington	Washington	98042	47.354838	-122.121185
415	2150 Deming Way	Middleton	Wisconsin	53562-5507	43.100195	-89.522751
416	950 Port Washington Rd	Grafton	Wisconsin	53024-9201	43.324691	-87.921615

417 rows × 6 columns

In [8]:
```
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
```

In [9]:
```
#https://basemaptutorial.readthedocs.io/en/latest/plotting_data.html

m = Basemap(projection='mill',
            lat_0=39, lon_0=-100,
           llcrnrlon=-165,llcrnrlat=15,urcrnrlon=-62,urcrnrlat=70,)

#good god this is a PITA
#AK 71, -164
#FL 22, -78
#ME 46, -63
#HI 18, -155

m.drawmapboundary(fill_color='aqua')
m.fillcontinents(color='coral',lake_color='aqua')
m.drawcoastlines()

lons = costco['Longitude']
lats = costco['Latitude']

x, y = m(lons, lats)

m.scatter(x, y, marker='o',color='m',zorder=5)
plt.title("Python - Spatial Chart: Costco Locations in the USA")
plt.show()
```

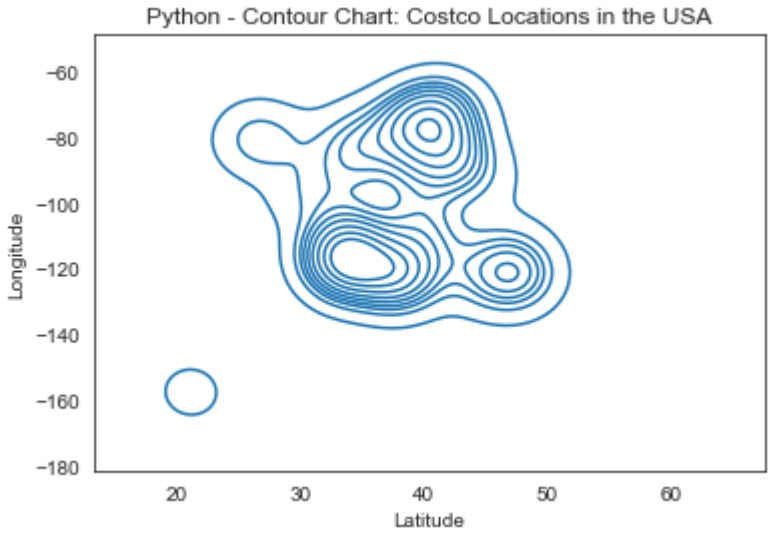


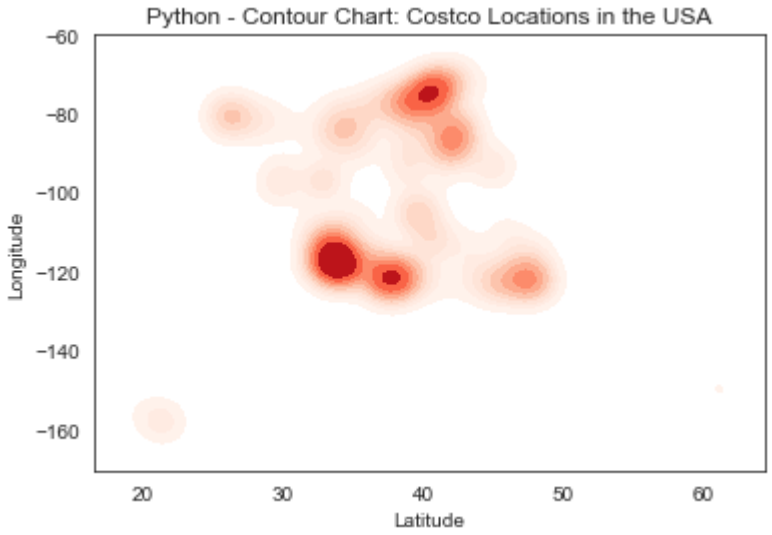
Contour Plot

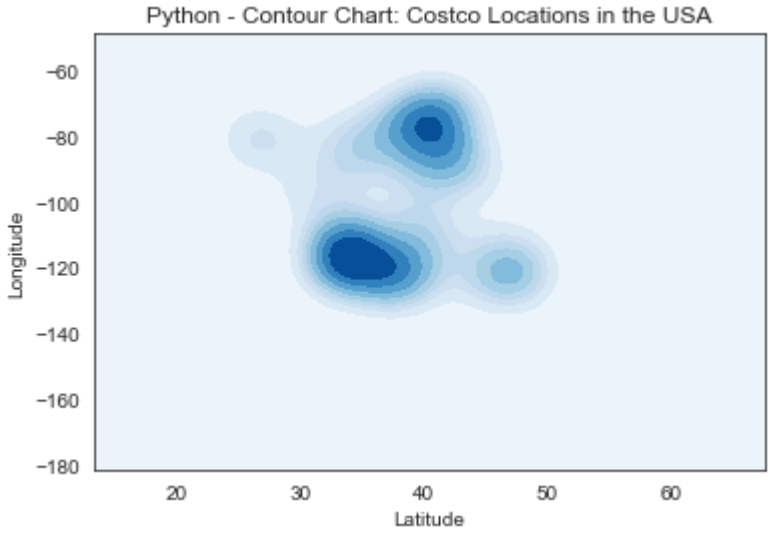
In [10]:
```
costco.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 417 entries, 0 to 416
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   Address    417 non-null    object 
 1   City       417 non-null    object 
 2   State      417 non-null    object 
 3   Zip Code   417 non-null    object 
 4   Latitude   417 non-null    float64
 5   Longitude  417 non-null    float64
dtypes: float64(2), object(4)
memory usage: 19.7+ KB
```

In [11]:
```
# set seaborn style
sns.set_style("white")

# Basic 2D density plot
sns.kdeplot(x=costco.Latitude, y=costco.Longitude)
plt.title("Python - Contour Chart: Costco Locations in the USA")
plt.show()

# Custom the color, add shade and bandwidth
sns.kdeplot(x=costco.Latitude, y=costco.Longitude, cmap="Reds", shade=True, bw_ad
plt.title("Python - Contour Chart: Costco Locations in the USA")
plt.show()

# Add thresh parameter
sns.kdeplot(x=costco.Latitude, y=costco.Longitude, cmap="Blues", shade=True, thre
plt.title("Python - Contour Chart: Costco Locations in the USA")
plt.show()
```






Python - Contour Chart: Costco Locations in the USA
−60
−80
−100
−120
−140
−160
−180
Longitude
20
30
40
50
60
Latitude

Week 9 & 10

Code

Hide

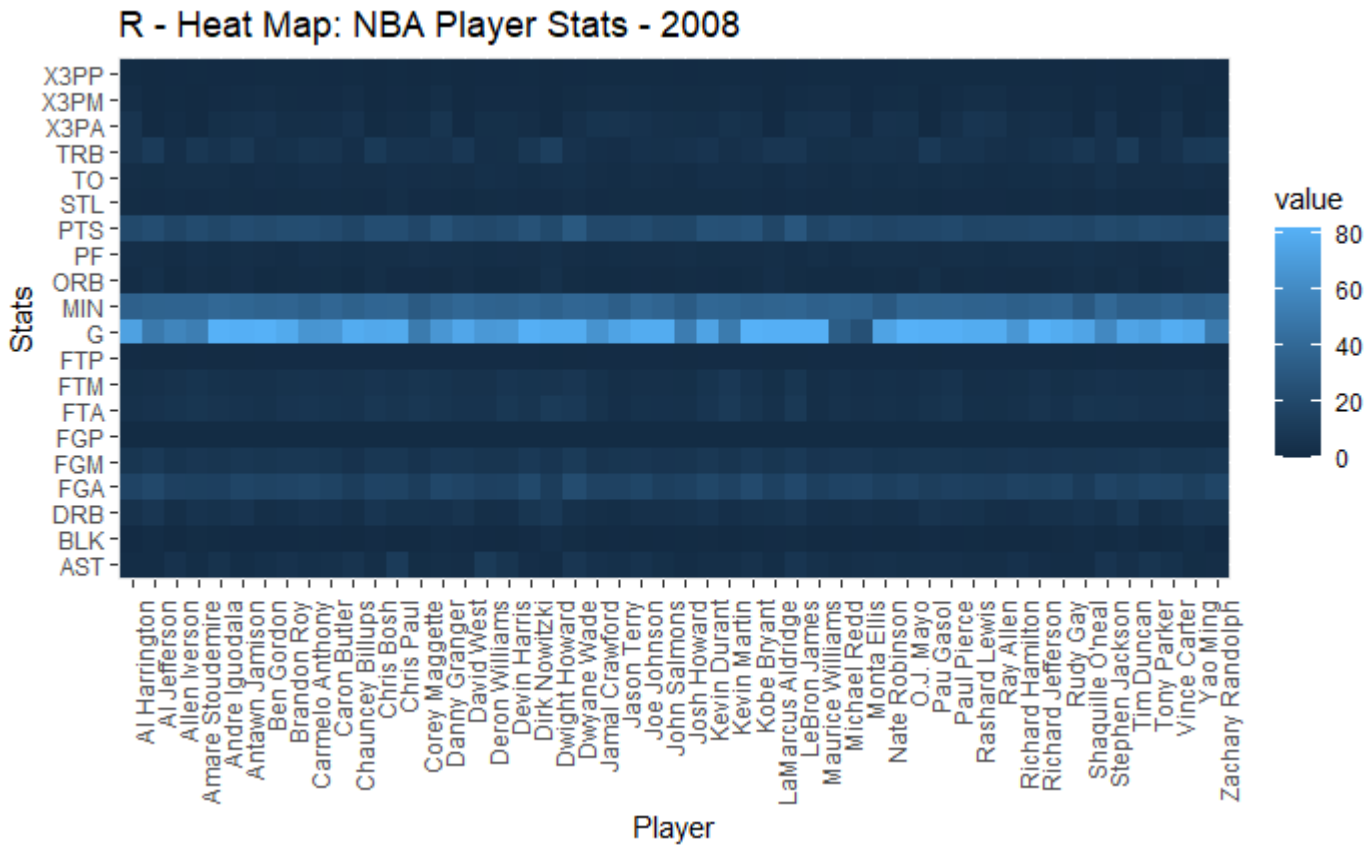
```
#load libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(tidyverse)
library(hrbrthemes)
library(pivottabler)
library(areaplot)
```

Hide

```
#import data
nba = read.csv("C:\\Users\\longr\\Documents\\DSC 640\\Week 9 & 10\\5.2 Exercises\\ppg2008.csv",h
eader=TRUE, row.names="Name")
costco = read.csv("C:\\Users\\longr\\Documents\\DSC 640\\Week 9 & 10\\5.2 Exercises\\costcos-geo
coded.csv")
```

Hide

```
#Heat map
nbaheat <- nba%>%
  rownames_to_column() %>%
  gather(colname, value, -rowname)

ggplot(nbaheat, aes(x = rowname, y = colname, fill = value)) +
  geom_tile()+xlab("Player")+ylab("Stats")+
  ggtitle("R - Heat Map: NBA Player Stats - 2008")+
  theme(axis.text.x = element_text(angle=90, hjust=1))
```



Hide

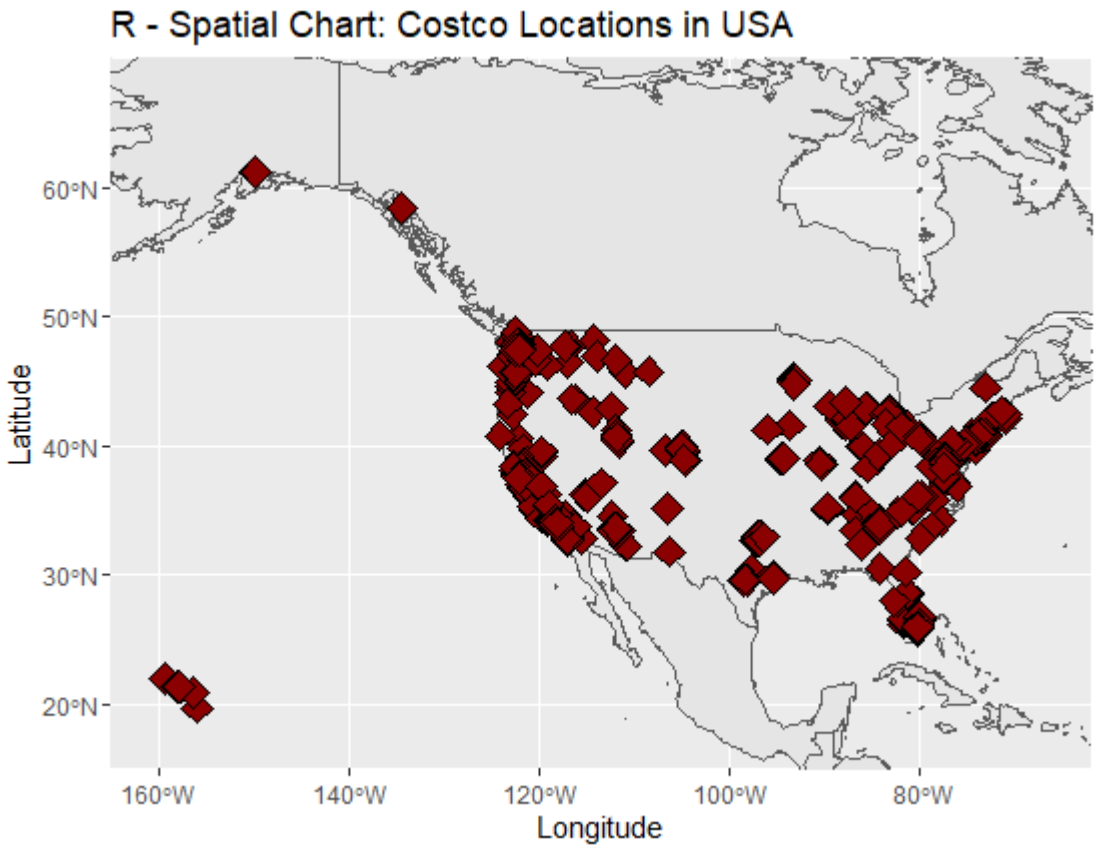
```
#Spatial Chart
#https://r-charts.com/spatial/bubble-map/
#https://r-spatial.org/r/2018/10/25/ggplot2-sf-2.html
```

Hide

```
library("rnaturalearth")
library("rnaturalearthdata")

world <- ne_countries(scale = "medium", returnclass = "sf")
```

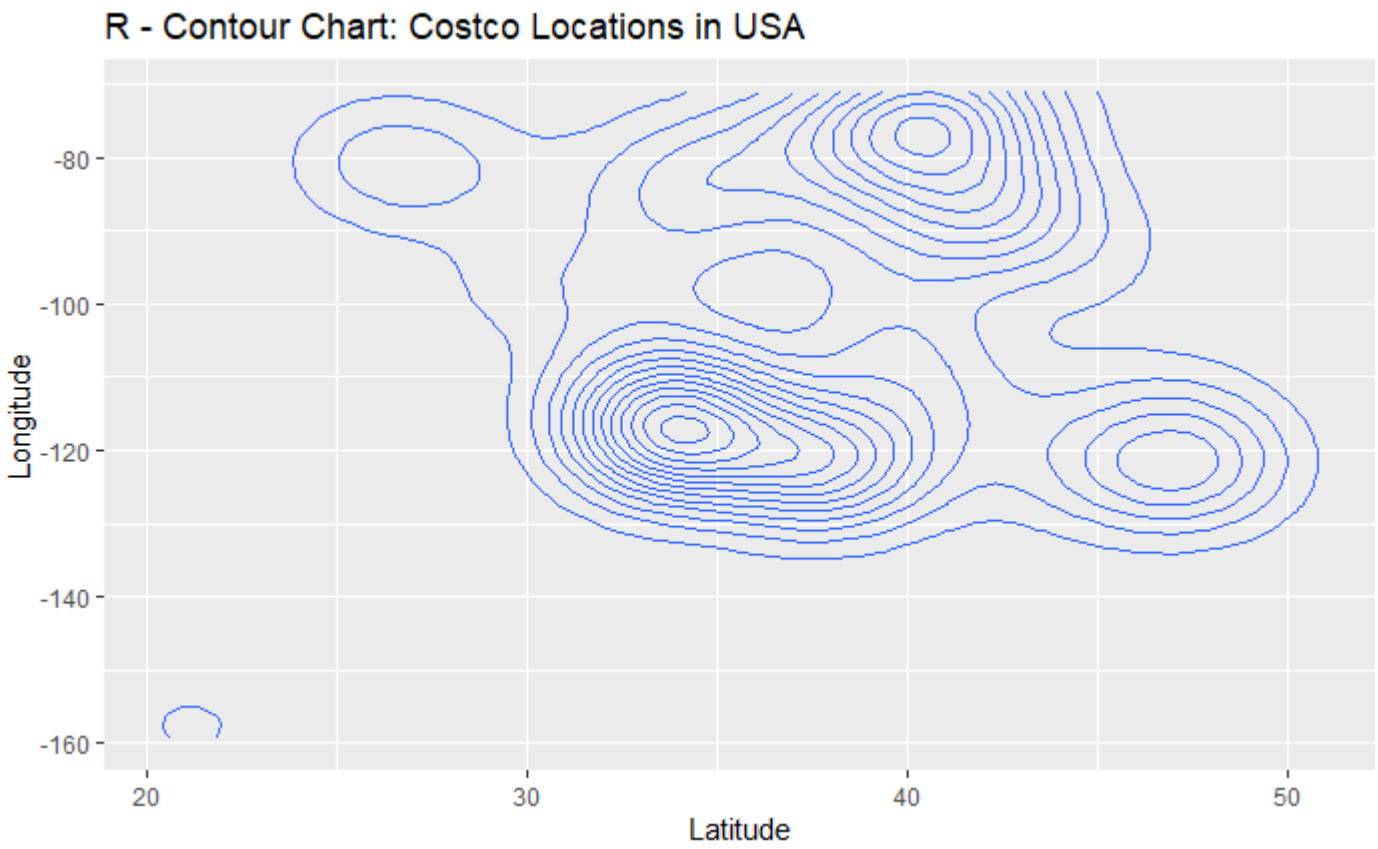
Hide

```
ggplot(data = world) +
    geom_sf() +
    geom_point(data = costco, aes(x = Longitude, y = Latitude), size = 4, shape = 23, fill = "da
rkred") +
    coord_sf(xlim = c(-62, -165), ylim = c(15, 70), expand = FALSE)+
    ggtitle("R - Spatial Chart: Costco Locations in USA")
```



Hide

```
#Contour Chart

ggplot(costco, aes(x = Latitude, y = Longitude)) +
  geom_density_2d()+
  ggtitle("R - Contour Chart: Costco Locations in USA")
```


R - Contour Chart: Costco Locations in USA
-80
-100
-120
-140
-160
Longitude
20
30
40
50
Latitude