Power BI - Heat Map

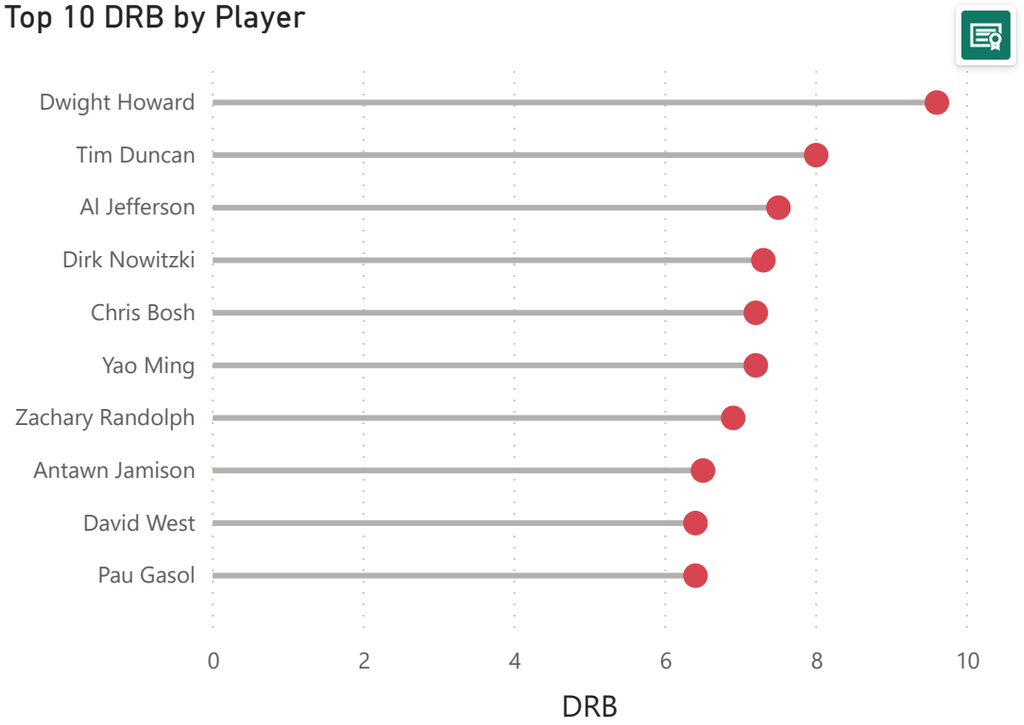
TV Sizes Sold by Year

size	2001	2002	2003	2004	2005	2006	2007	2008	2009
24									
32									
40									
42									
43									
50									
55									
60									
65									

Power BI - Spatial Map

Costco Store Locations

Power BI - Lollipop



Campbell640Week9-10

November 3, 2023

[1]:
```
# Load libraries
import pandas as pd
import numpy as np
# import data visualization libraries
import seaborn as sns
import matplotlib.pyplot as plt
# Plotly
import folium
# Image Export
import io
from PIL import Image
from IPython.display import Image
```

[2]:
```
costco = pd.read_csv('data sources/ex5-2/costcos-geocoded.csv')
costco.shape
```

[2]:
```
(417, 6)
```

[3]:
```
ppg = pd.read_csv('data sources/ex5-2/ppg2008.csv')
ppg.shape
```

[3]:
```
(50, 21)
```

[4]:
```
tv = pd.read_csv('data sources/ex4-2/tv_sizes.txt', sep='\t')
tv.shape
```

[4]:
```
(764, 2)
```

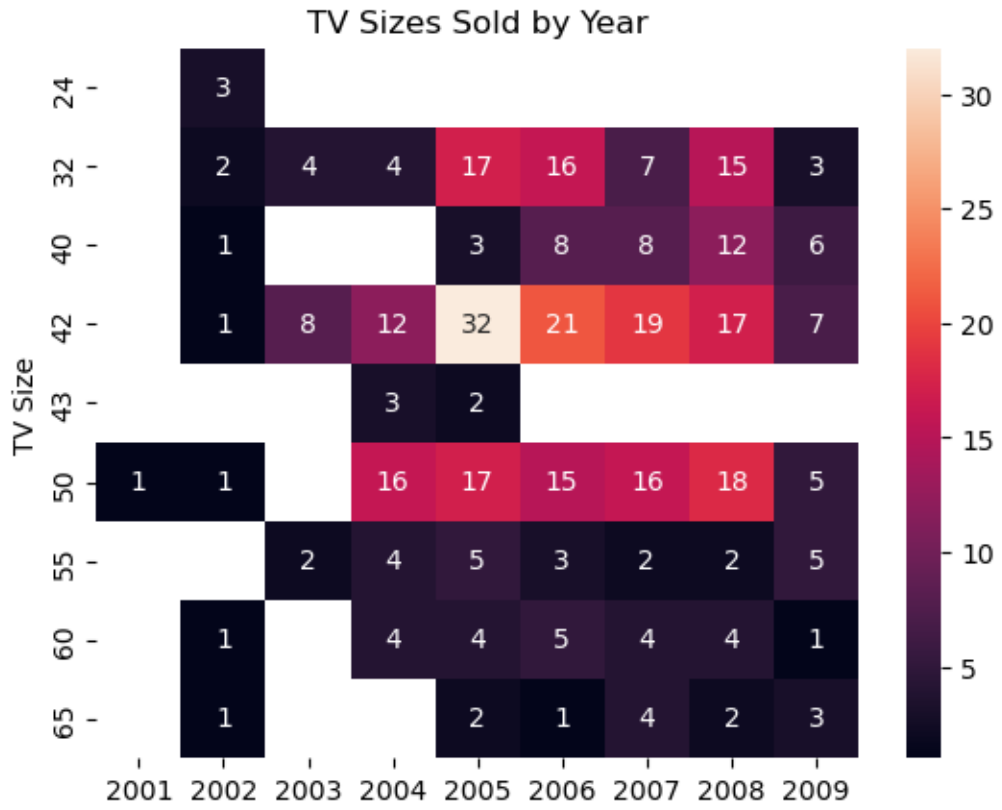
1 Python - Heat Map

[22]:
```
# Get count of tv by tv size
tv_size = tv.groupby(['year', 'size'])['size'].count().reset_index(name="count")
```

[43]:
```
# Filter to popular tv sizes
size = [24, 32, 40, 42, 43, 50, 55, 60, 65]
tv_filtered = tv_size[tv_size['size'].isin(size)R ]
```

[48]:
```
# Create pivot table
df_heatmap = tv_filtered.
  ↪pivot_table(values='count',index='size',columns='year',aggfunc=np.mean)
```

[52]:
```
# Display heatmap
sns.heatmap(df_heatmap,annot=True)
plt.title('TV Sizes Sold by Year')
plt.xlabel('')
plt.ylabel('TV Size')
plt.show()
```



2 Python - Spatial Map

[5]:
```
# Find States
costco_states = costco.groupby('State', as_index=False).mean()
```

[6]:
```
# Get count of costcos per state
costco_counts = costco.value_counts(subset=['State']).reset_index(name='counts')
```

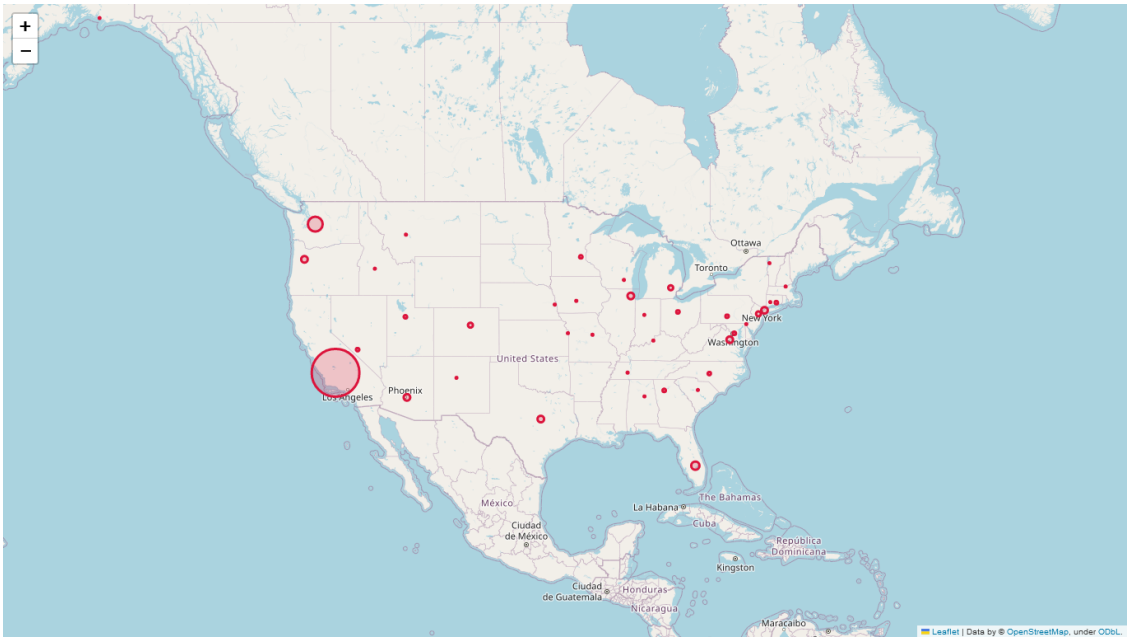
[8]:
```
# Merge
costco_merged = costco_states.merge(costco_counts,how="left",on="State")
```

[9]:
```
# initialize the map and store it in a m object
m = folium.Map(location = [40, -95],
               zoom_start = 4)

# add marker one by one on the map
for state in costco_merged.itertuples():
    folium.Circle(
        location=[state.Latitude, state.Longitude],
        popup='%s (%.1f)' % (state.State, state.counts),
        radius=state.counts * 2000,
        color='crimson',
        fill=True,
        fill_color='crimson'
    ).add_to(m)
```

[12]:
```
img_data = m._to_png(5)
img = Image.open(io.BytesIO(img_data))
img.save('image.png')
```

[15]:
```
Image("image.png")
```

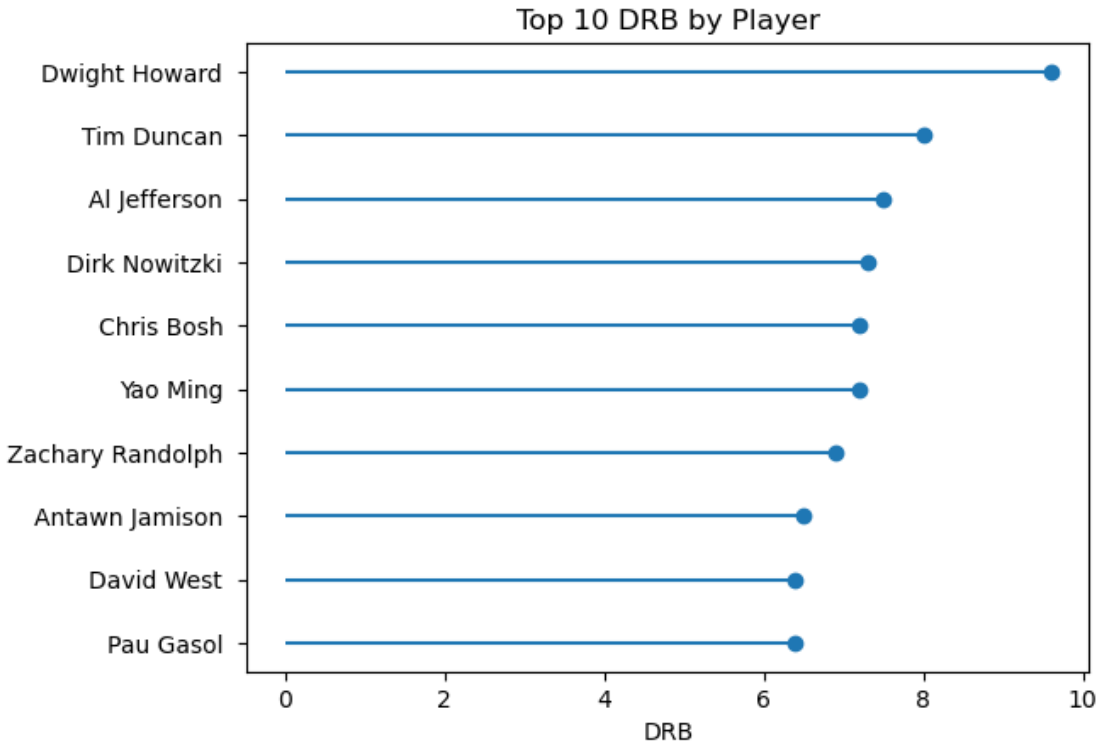
[15]:



3 Python - Lollipop

[114]:
```
# Get top 10 players by DRB
top_ppg = ppg.sort_values(by='DRB', ascending=False)
top_ppg = top_ppg[:10].sort_values(by='DRB')
```

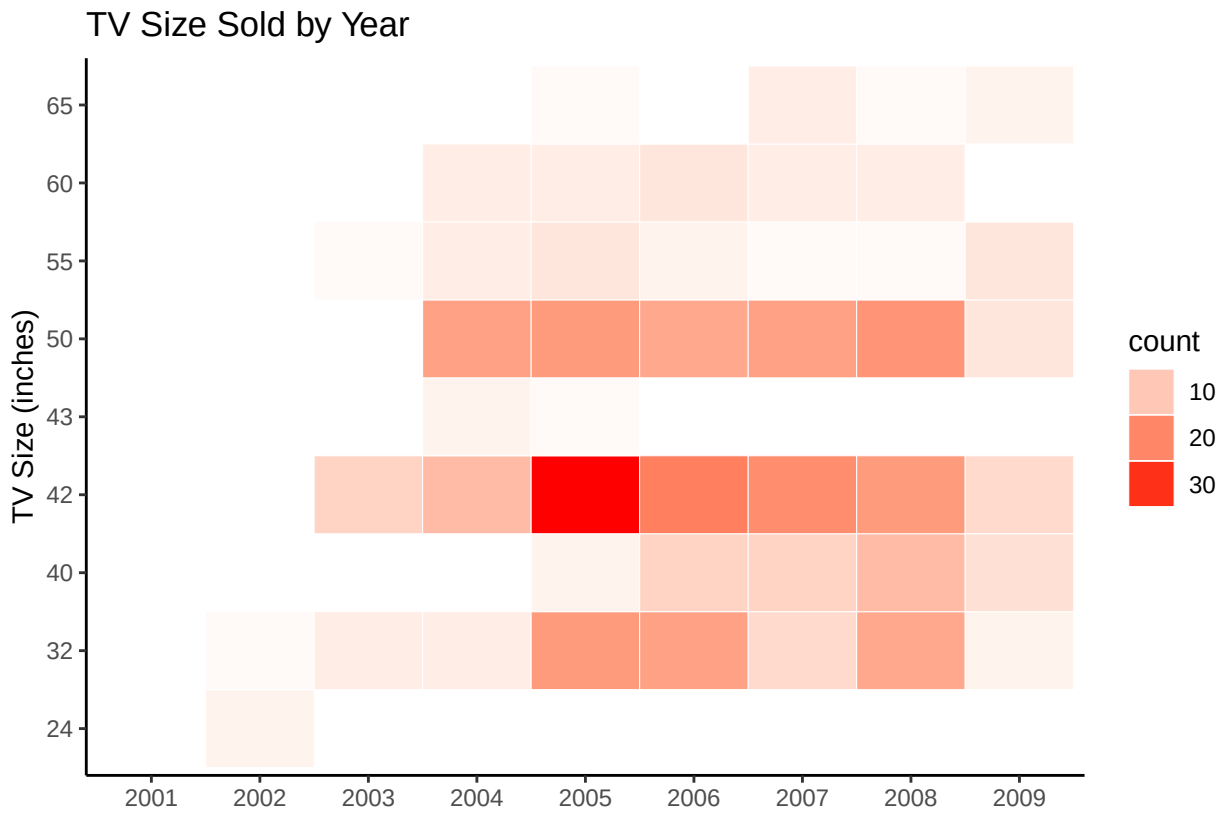
[116]:
```
# Draw lollipop chart
plt.hlines(top_ppg['Name  '], xmin=0,
            xmax=top_ppg['DRB'])

# drawing the markers (circle)
plt.plot(top_ppg['DRB'], top_ppg['Name  '], "o")
plt.xlabel('DRB')
plt.title('Top 10 DRB by Player')
plt.show()
```



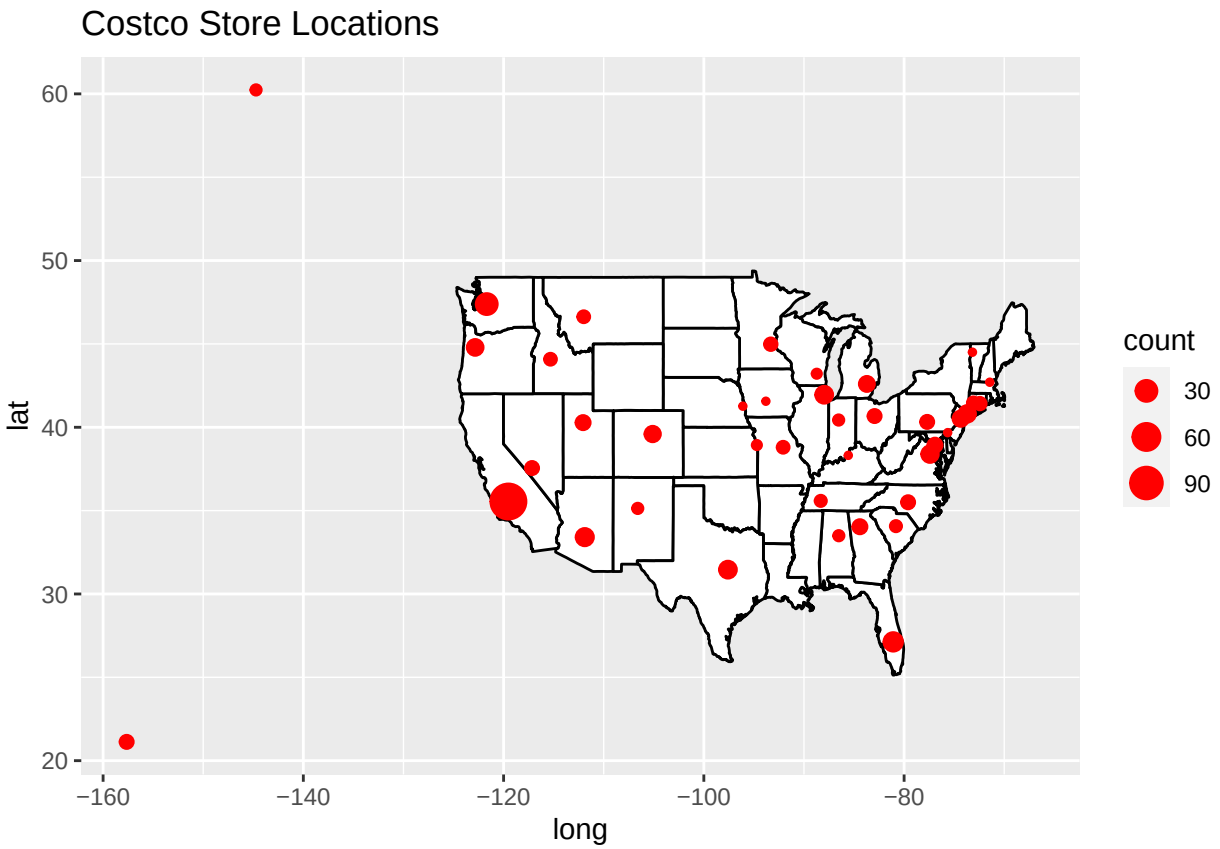
Campbell640Week9-10

2023-10-24

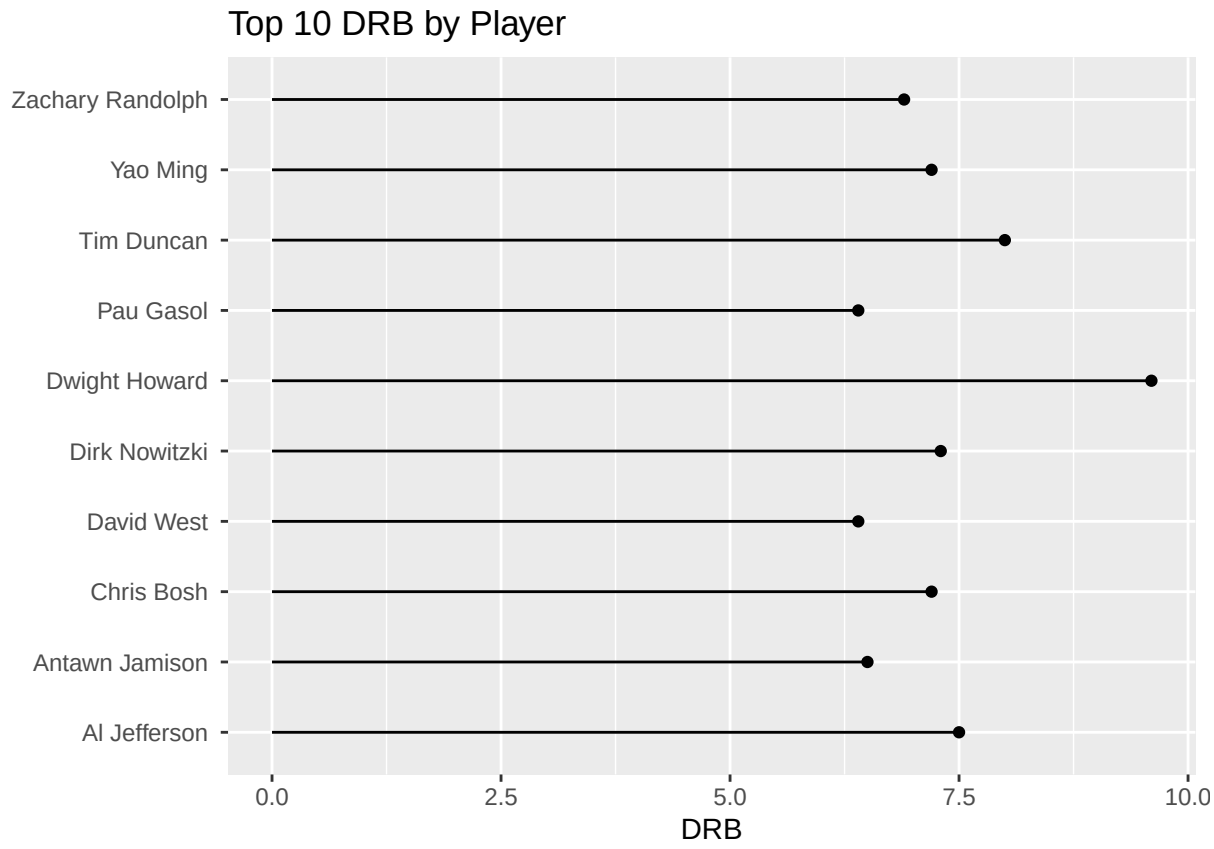
R - Heat Map



R - Spatial Map



R - Lollipop



Code Repository

```
## Set the working directory to the root of your DSC 640 directory
setwd("C:/Users/jcamp/Documents/DSC640/Assignments/data sources")

# Load libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(reshape)
library(maps)
# Load file
costco <- read.csv("ex5-2/costcos-geocoded.csv")
ppg <- read.csv("ex5-2/ppg2008.csv")
tv <- read.table(file = "ex4-2/tv_sizes.txt", header = TRUE)
# Group expenditures by category
tv_count <- tv %>%
            group_by(year, size) %>%
            count(size) %>%
            filter_at(vars(size), any_vars(. %in% c(24,32,40,42,43,50,55,60,65))) %>%
            replace(is.na(.), 0)


ggplot(tv_count, aes(y = as.factor(size), x = as.factor(year), fill = n)) +
```

```
  geom_tile(color="white") +
  scale_fill_gradient(low = "white", high = "red") +
  theme_classic()+
  ggtitle("TV Size Sold by Year") +
  xlab("") +
  ylab("TV Size (inches)") +
  guides(fill=guide_legend(title="count"))
# Get states map info
states <- map_data("state")

# Get count of Costcos
costco_state <- costco %>%
                  group_by(State,.drop = FALSE) %>%
                  summarize(lat = mean(Latitude),
                            lon = mean(Longitude),
                            count = n())
# Plot on states map
ggplot(data = states, mapping = aes(x = long, y= lat, group = group)) +
  geom_polygon(fill = "white", color = "black") +
  geom_point(data = costco_state, color = "red",
             aes(x = lon, y = lat, group = NULL, size = count))+
  ggtitle("Costco Store Locations")
top10 <- ppg[order(ppg$DRB, decreasing = TRUE), ][1:10,]

ggplot(top10, aes(x = DRB, y = Name)) +
  geom_segment(aes(x = 0, y = Name, xend = DRB, yend = Name)) +
  geom_point()+
  ggtitle("Top 10 DRB by Player") +
  xlab("DRB") +
  ylab("")
```