College Basketball: Season Statistics Visualizations

Granger Moch, Will Murphy

October 1, 2014

Contents

Data

Using the seasonstatistics dataset, subsets will be created to visualize disinct conferences, players, and teams.

Initiate needed libraries

```
library(reshape2)
library(ggplot2)
```

Create a subset of West Coast Conference players/teams

```
westcoast=subset(Seasonstatistics, Conference== "West Coast")
```

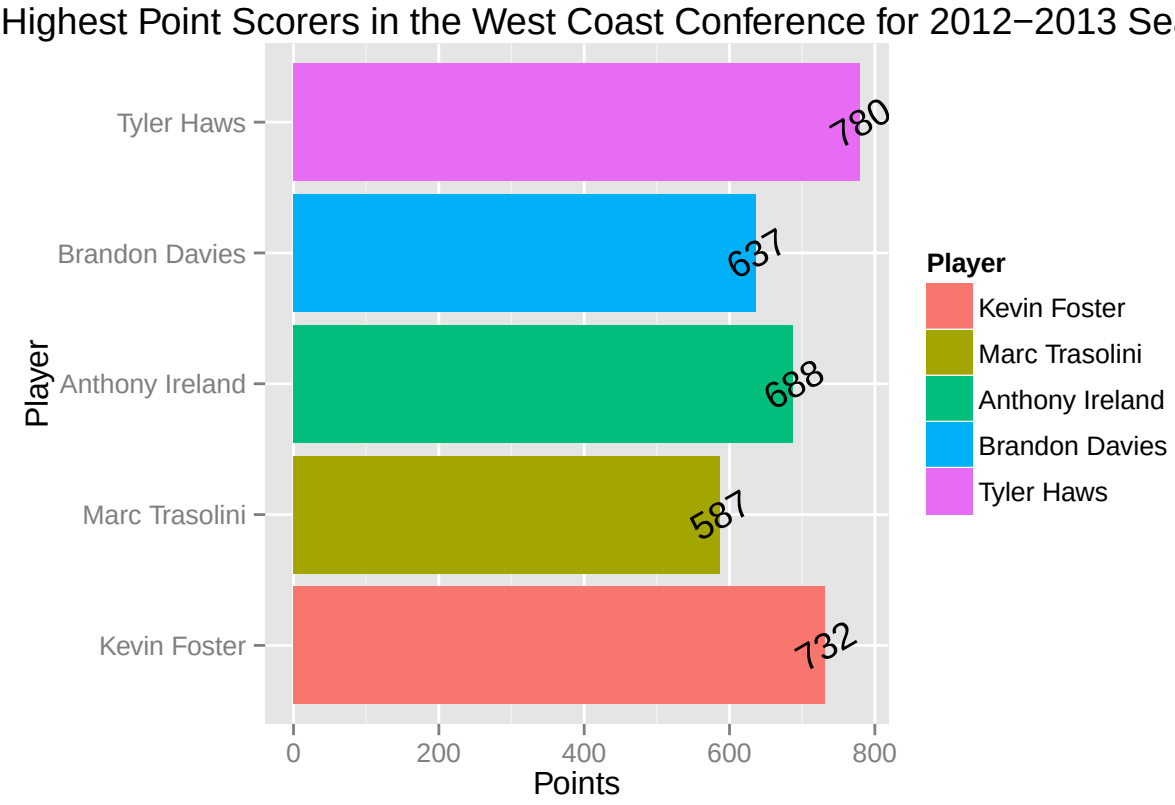
Make a subset of top five highest scorers from West Coast conference for 2012-2013.

```
westcoast2012=subset(westcoast,Year== "2012-2013")
topfive=subset(westcoast2012,Points >580)
```

The data from this subset can now be plotted.

Graph Showing the Top 5 Highest Point Scorers in WCC Conference for 2012-2013 Season

```
t5=ggplot(topfive,aes(x=Player, y=Points, fill=Player, label=Points)) +geom_bar(stat="identity")

t5 +ggtitle("Top 5 Highest Point Scorers in the West Coast Conference for 2012-2013 Season") +
geom_text(angle=30) +coord_flip()
```



While this graph's data density is adequate, I felt that it could display more about just how effective these top five scorers are when it comes to shooting (2 point) field goals. So, I would like to create a stacked bar plot, with the fill showing the the made field goal attempts vs missed field goal attempts.

However, the dataset provided by ESPN does not have a statistic for field goals missed. . . only field goals made and attempted. Having these two variables is sufficient enough to formulate the field goals missed column. The column will first be added to the dataset and then it will be populated with data.

add Field Goals Missed column to topfive dataset (currently null)

```
topfive["Field.Goals.Missed"]=NA
```

add data to Field Goals Missed column

```
topfive$Field.Goals.Missed= (topfive$Field.Goals.Attempted - topfive$Field.Goals.Made)
```

Create a subset holding these three relevant variables

```
madeMiss <- subset(topfive, Conference=="West Coast",
select=c(Player,Field.Goals.Made, Field.Goals.Missed))
```

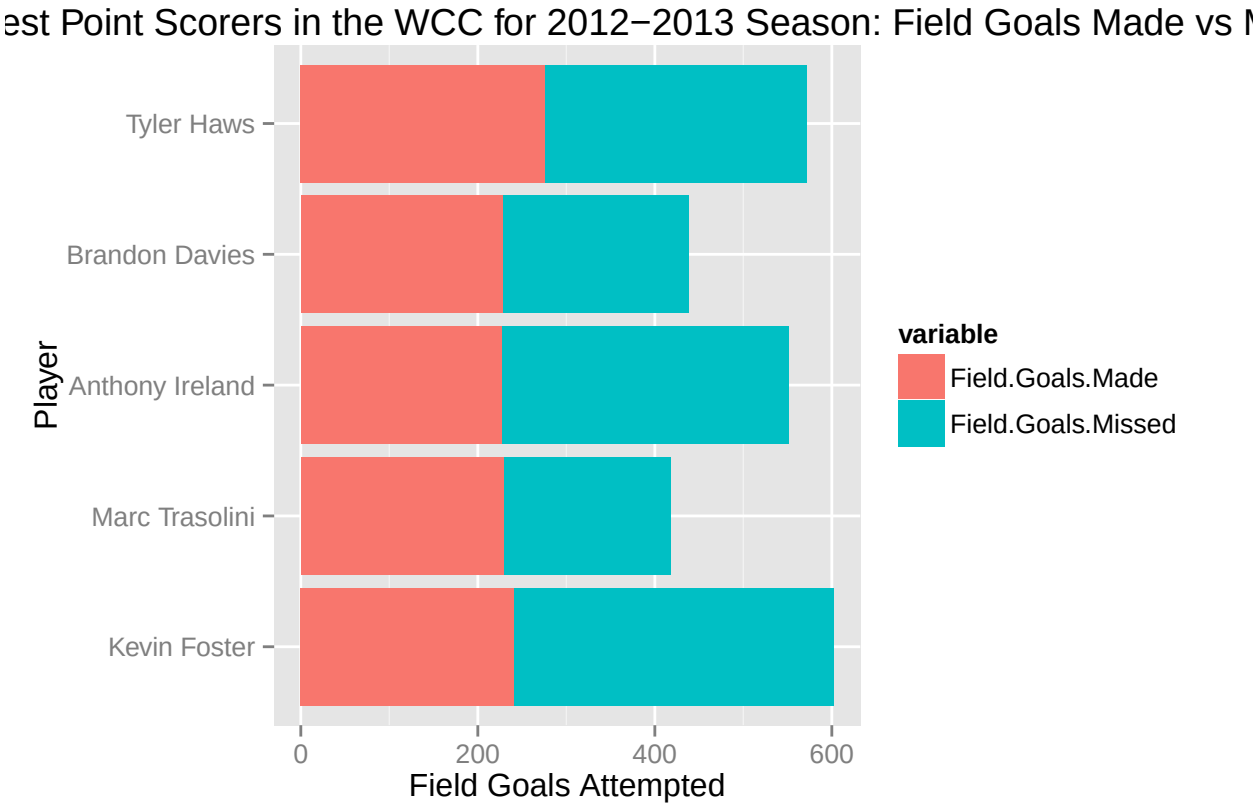
The fill aesthetic needs a variable by which it should fill the geoms, bars in this case.You can't use a dataframe for a fill, which is what I had initially tried. What is needed is to reshape the dataframe such that field goals made and field goals attempted become variables in one column and another column has the values for these. The third column is the column of player names.

Reshape topfive dataset using Player as id

```
topfivemelt=melt(madeMiss,id="Player") #reshaping dummydata to format needed... see the structure of du
```

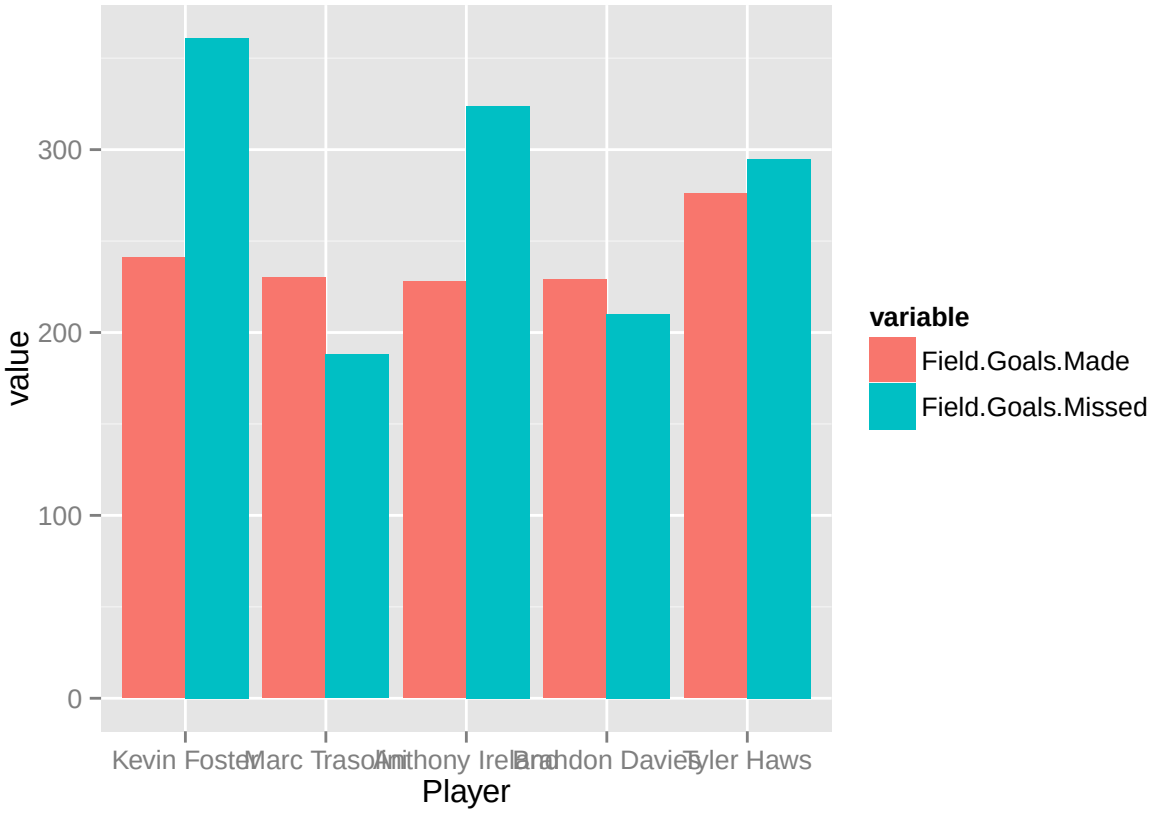
Stacked Bar Plot of Top Five Highest Point Scorers in the WCC for 2012-2013 Season: Field Goals Made vs Missed

```
ggplot(topfivemelt,aes(x=Player,y=value, fill=variable))+geom_bar(stat="identity") +coord_flip() +ylab(
```



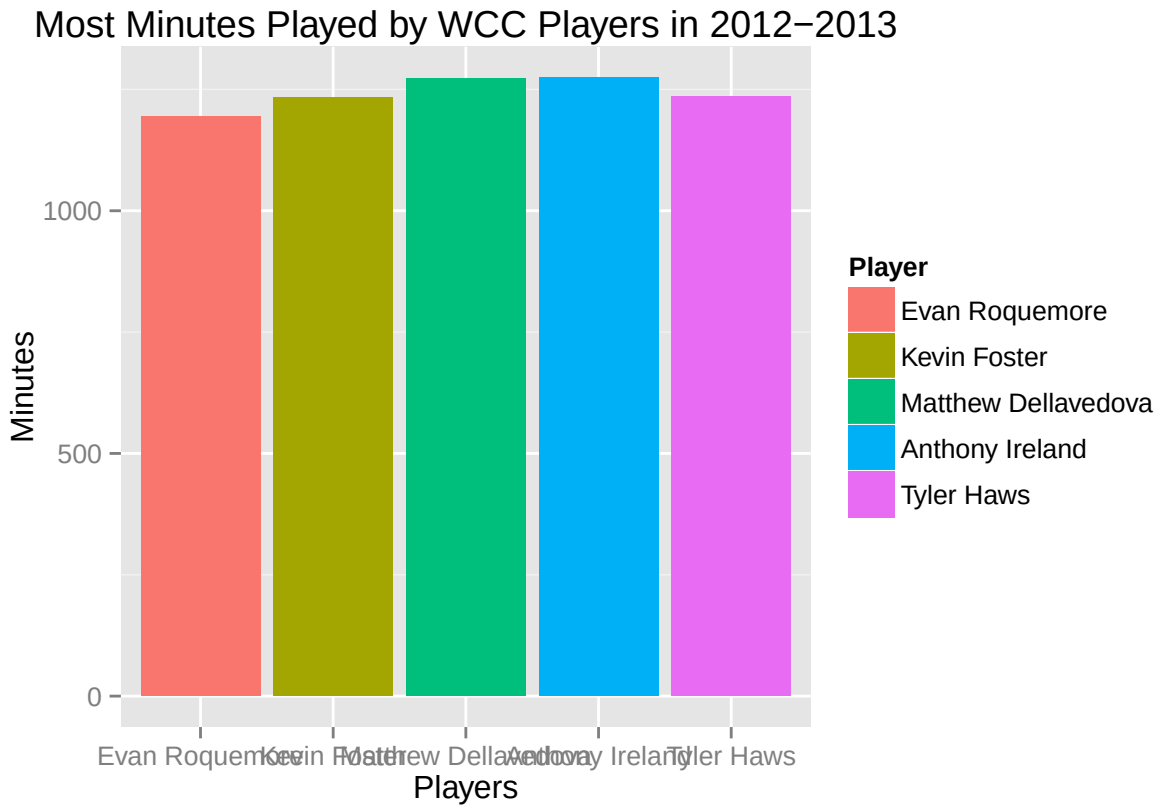
Regular side-by-side bar plot showing same data

```
ggplot(topfivemelt,aes(x=Player,y=value,fill=variable))+geom_bar(stat="identity",position="dodge") # st
```



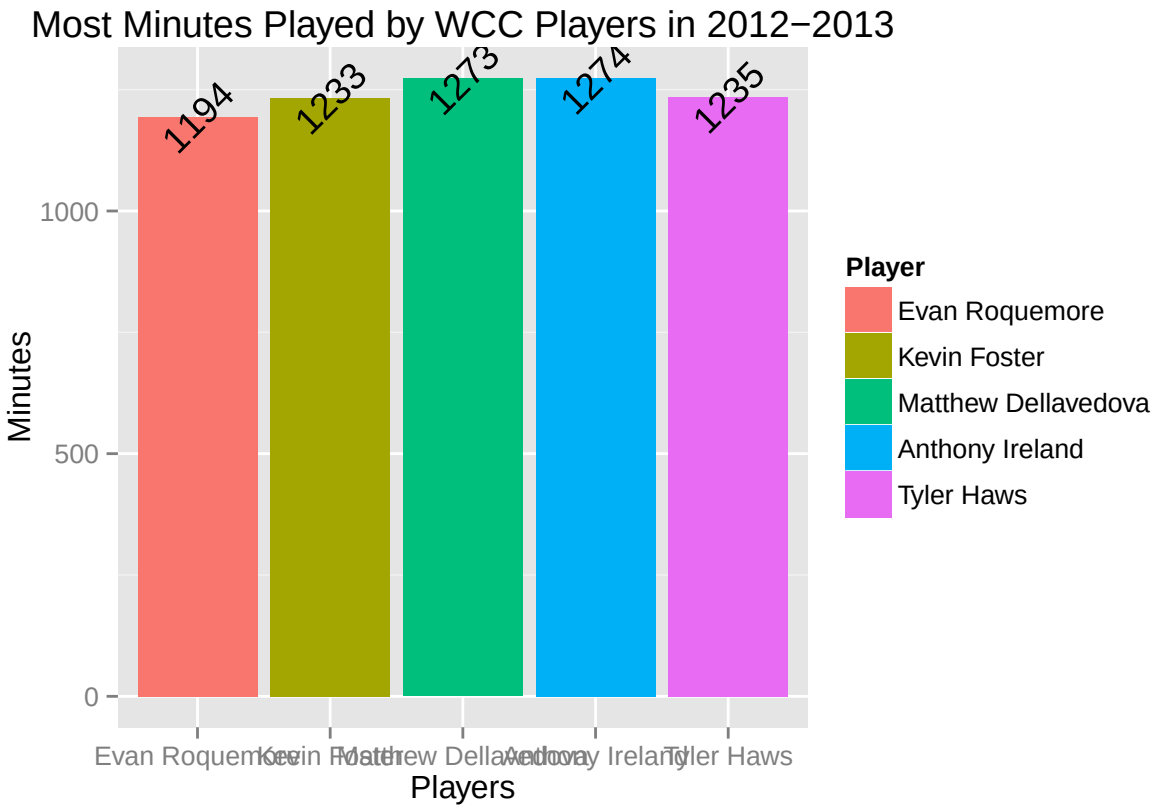
Most Minutes Played by Players in WCC for 2012-2013 Season

```
mostminutes=subset(westcoast2012, Minutes > 1190)

mm=ggplot(mostminutes,aes(x=mostminutes$Player, y=mostminutes$Minutes, fill=Player, label=Minutes))+geo

mm
```



Improved readability of data

```
mm+geom_text(angle=45)
```



Top Five Highest Points Per Minute for WCC Players in 2012-2013

```
#add Points.Per.Minute column to westcoast2012 dataset
westcoast2012["Points.Per.Minute"]=NA

#populate new column
westcoast2012$Points.Per.Minute=(westcoast2012$Points/westcoast2012$Minute)


westcoast2012["Points.Per.Minute.Rounded"]=NA

#round minutes to nearest hundredth
westcoast2012$Points.Per.Minute.Rounded= format(round(westcoast2012$Points.Per.Minute,3))

#create subset that removes non integers/ infinite values and places contraint on points per minute to
ppm=subset(westcoast2012,westcoast2012$Points.Per.Minute.Rounded !="NaN" & westcoast2012$Points.Per.Minu

ggplot(ppm, aes(x=reorder(Player, Points.Per.Minute),y=Points.Per.Minute, fill=Player, label=Points.Per
```