

Project 2

John Rollman

October 14, 2020

Introduction

I will write stuff here about the bike data. there are variables about the day such as year, holiday, weekday, working day and the weather. The purpose of this project is to predict the count of bikes rented on a particular day given some covariates about the day.

Packages Used

```
library(tidyverse)
library(caret)
library(knitr)
library(gbm)
library(randomForest)
library(DT)
library(corrplot)
library(cowplot)
library(gridExtra)
library(ggpubr)
```

Data Pulling and Manipulation

```
bikeData.full <- read.csv("day.csv")
bikeData <- bikeData.full %>% as_tibble() %>% select(!casual & !registered) %>% filter(weekday==1) ##Ad
```

Exploratory Summary

```
#Scatter Plots
tempPlot <- ggplot(bikeData,aes(x=temp,y=cnt)) +
  geom_point(aes(color = season)) +
  geom_smooth(method=lm, col="Green") +
  ggtitle("Temperature vs Daily Bike Rentals") +
  labs(x = "Normalized Temp in C", y = "Daily Bike Rental Count")

atempPlot <- ggplot(bikeData,aes(x=atemp,y=cnt)) +
  geom_point(aes(color = season)) +
  geom_smooth(method=lm, col="Green") +
  ggtitle("'Feeling' Temperature vs Daily Bike Rentals") +
  labs(x = "Normalized 'Feeling' Temp in C", y = "Daily Bike Rental Count")

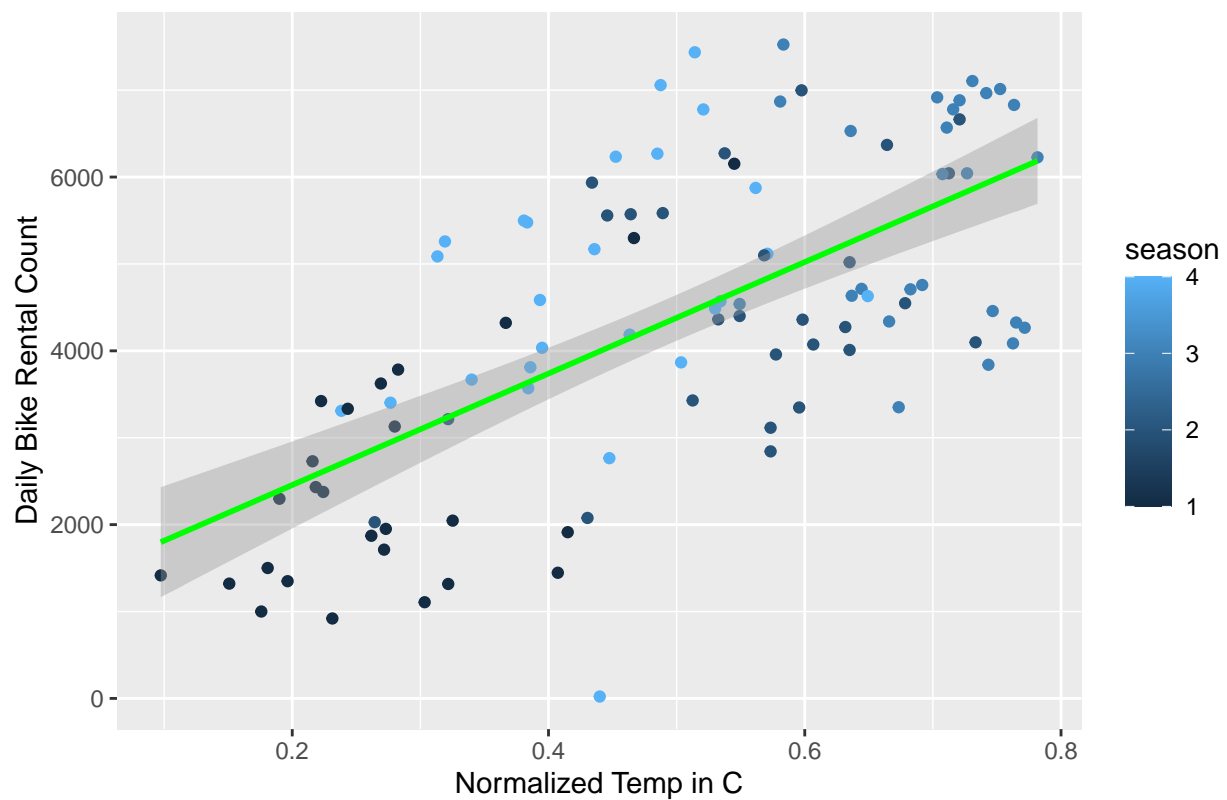
humPlot <- ggplot(bikeData,aes(x=hum,y=cnt)) +
```

```
geom_point(aes(color = season)) +
geom_smooth(method=lm, col="Green") +
ggtitle("Humidity vs Daily Bike Rentals") +
labs(x = "Normalized Humidity", y = "Daily Bike Rental Count")
```

```
wsPlot <- ggplot(bikeData,aes(x=windspeed,y=cnt)) +
geom_point(aes(color = season)) +
geom_smooth(method=lm, col="Green") +
ggtitle("Wind Speed vs Daily Bike Rentals") +
labs(x = "Normalized Wind Speed", y = "Daily Bike Rental Count")
```

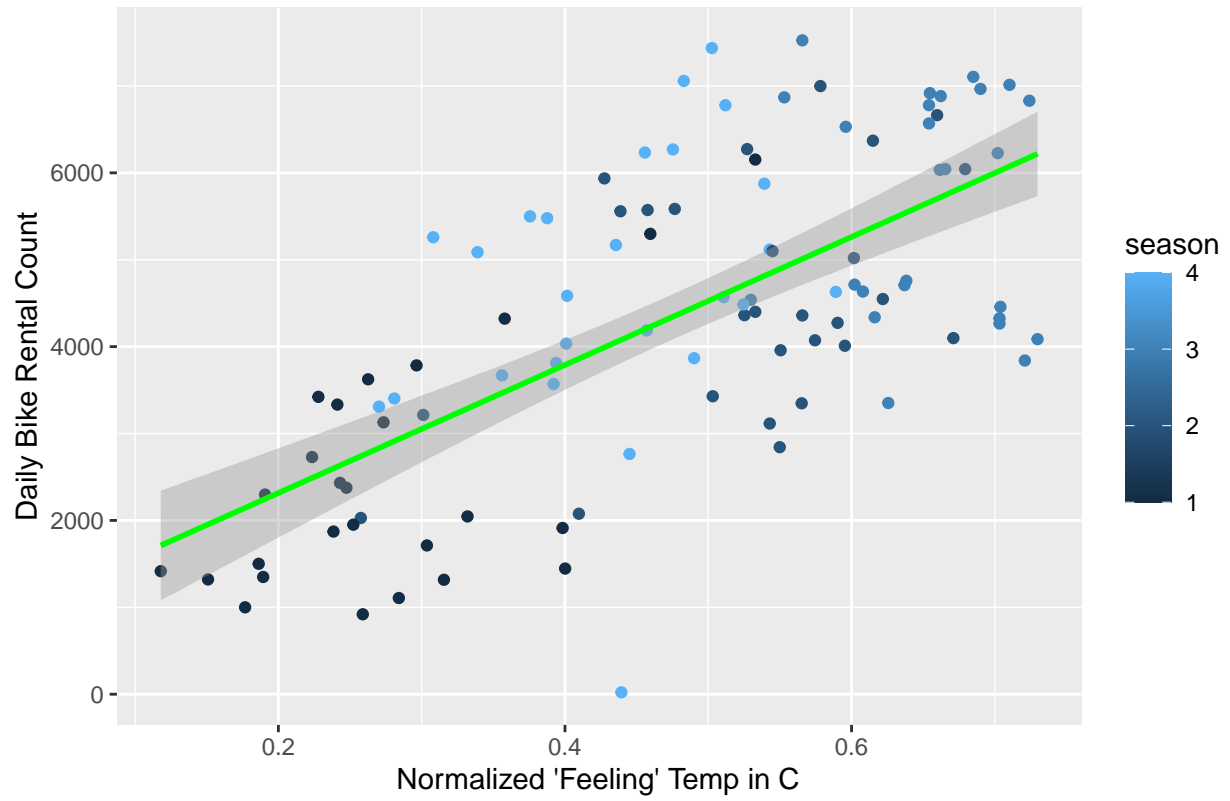
tempPlot

Temperature vs Daily Bike Rentals



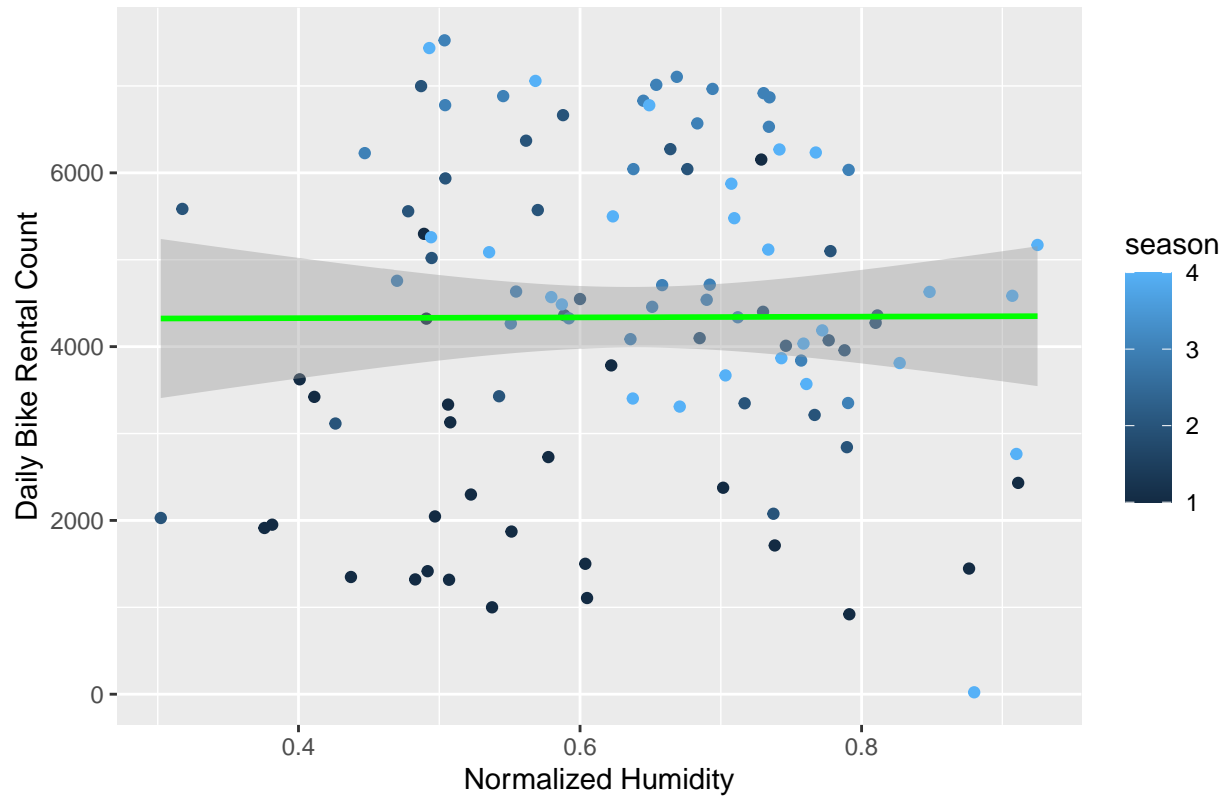
atempPlot

'Feeling' Temperature vs Daily Bike Rentals



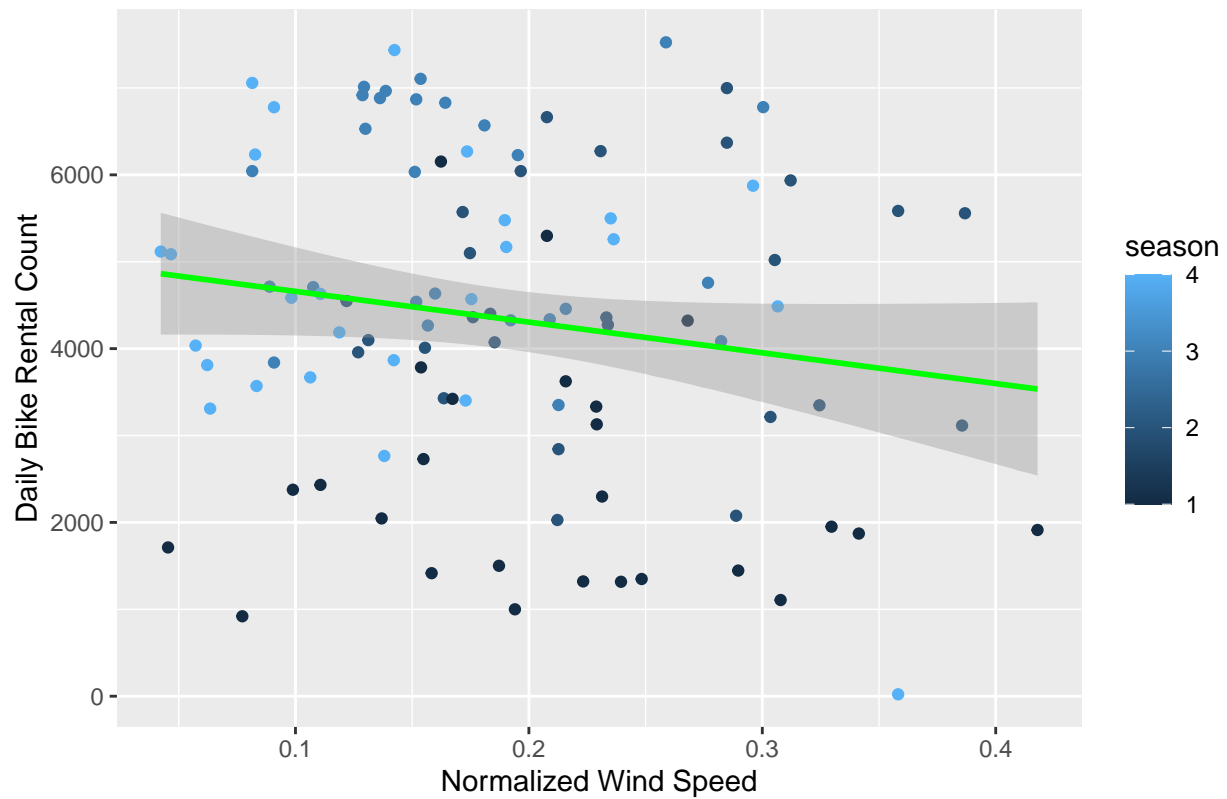
humPlot

Humidity vs Daily Bike Rentals



wsPlot

Wind Speed vs Daily Bike Rentals

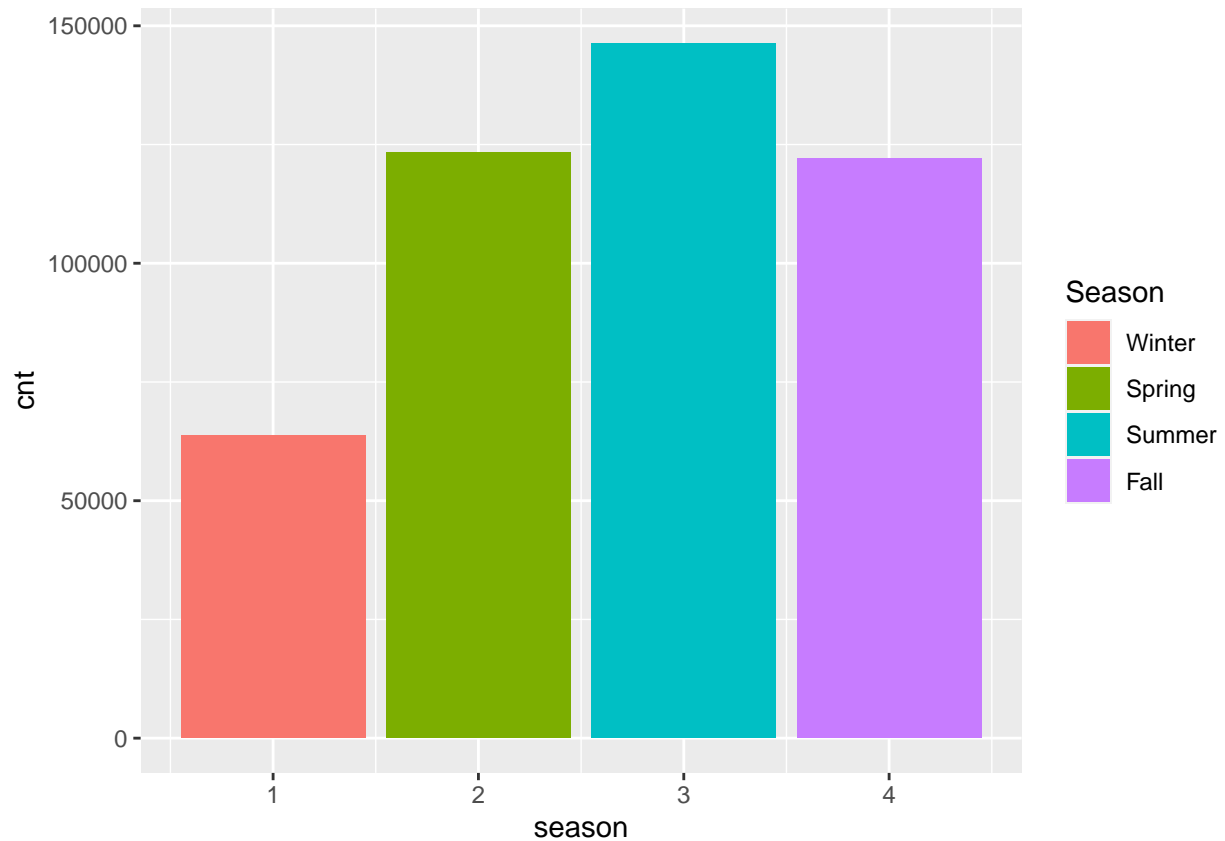


```
#plot_grid(tempPlot, atempPlot, humPlot, wsPlot)

#grid.arrange(tempPlot, atempPlot, humPlot, wsPlot)

#ggarrange(tempPlot, atempPlot, humPlot, wsPlot, common.legend = T, legend = "right")

#Categorical Summaries
ggplot(bikeData, aes(x = season, y = cnt)) +
  geom_bar(stat = "identity", aes(fill = as.factor(season))) +
  scale_fill_discrete(name = "Season", labels=c("Winter", "Spring", "Summer", "Fall"))
```



Model Building

Data Splitting

```
set.seed(558)
DataIndex <- createDataPartition(bikeData$cnt, p = .7, list = F)
bikeTrain <- bikeData[DataIndex,]
bikeTest <- bikeData[-DataIndex,]
```