WU #8 - Interaction & Indicators

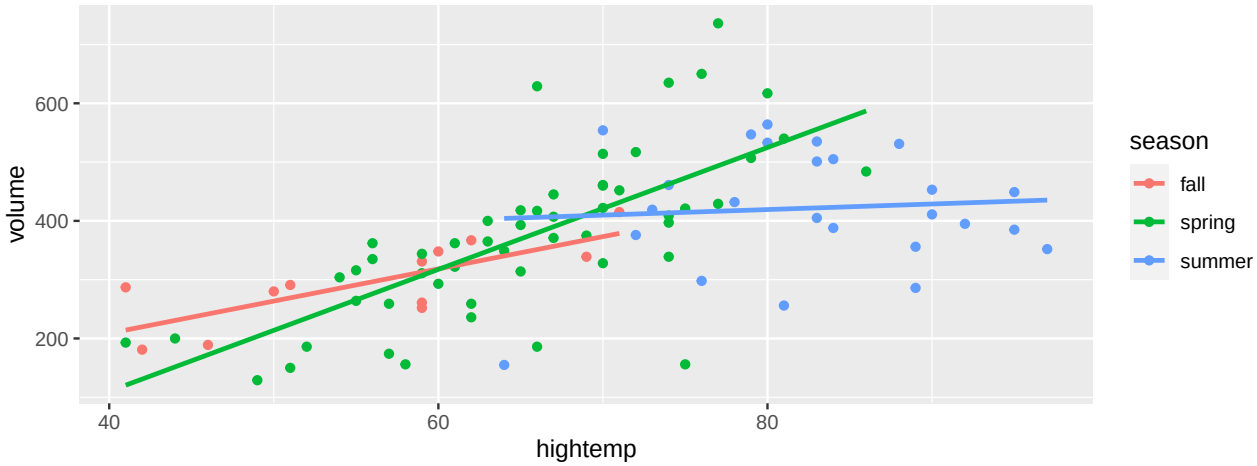
Math 158 - Jo Hardin

Tuesday 2/15/2022

Name: _____

Names of people you worked with: _____

Consider the `RailTrail` data. The model below regresses `volume` of bicycle riders on `hightemp` and `season`.




```
RailTrail %>%
  lm(volume ~ hightemp * season, data = .) %>%
  tidy()
```

```
## # A tibble: 6 x 5
##   term                  estimate std.error statistic p.value
##   <chr>                    <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)             -10.5     167.     -0.0631  0.950
## 2 hightemp                  5.48      2.95    1.86    0.0667
## 3 seasonspring           -294.      190.     -1.54    0.126
## 4 seasonsummer            354.      255.      1.39    0.169
## 5 hightemp:seasonspring     4.88      3.26    1.50    0.138
## 6 hightemp:seasonsummer    -4.54      3.75   -1.21    0.230
```

Write down three separate models (each of spring, summer, and fall) for how `hightemp` predicts `volume` separately depending on the season. The three answers will each look something like:

$$\text{winter: } \widehat{\texttt{volume}} = 47 + 0.47 \cdot \texttt{hightemp}.$$

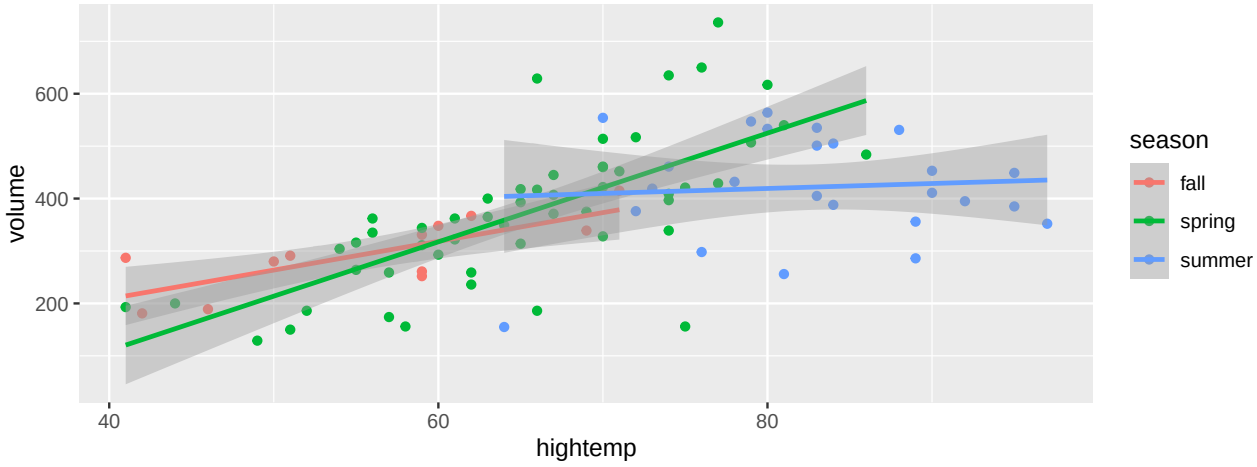
Solution:

fall: $\widehat{\texttt{volume}} = -10.53 + 5.48 \cdot \texttt{hightemp}$.

spring: $\widehat{\texttt{volume}} = (-10.53 - 293.95) + (5.48 + 4.88) \cdot \texttt{hightemp} = -304.48 + 10.36 \cdot \texttt{hightemp}$.

summer: $\widehat{\texttt{volume}} = (-10.53 + 354.18) + (5.48 + -4.54) \cdot \texttt{hightemp} = 343.65 + 0.94 \cdot \texttt{hightemp}$.

Note that the interaction term is not significant! The lack of significance can be seen if SE ribbons are added to the linear model.



Additionally, if the interaction is removed, the model shows that `hightemp` is significant, but `season` is not.

```
RailTrail %>%
  lm(volume ~ hightemp + season, data = .) %>%
  tidy()
```

```
## # A tibble: 4 x 5
##   term         estimate std.error statistic       p.value
##   <chr>           <dbl>     <dbl>     <dbl>         <dbl>
## 1 (Intercept)   -125.       71.7     -1.75  0.0841
## 2 hightemp         7.54      1.17     6.43  0.00000000692
## 3 seasonspring     5.13     34.3      0.150 0.881
## 4 seasonsummer   -76.8      47.7     -1.61  0.111
```

```
RailTrail %>%
  lm(volume ~ hightemp, data = .) %>%
  tidy()
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic       p.value
##   <chr>          <dbl>     <dbl>     <dbl>         <dbl>
## 1 (Intercept)   -17.1     59.4      -0.288 0.774
## 2 hightemp        5.70     0.848     6.72  0.00000000171
```