Homework 1

Kabir Snell

2022-04-20

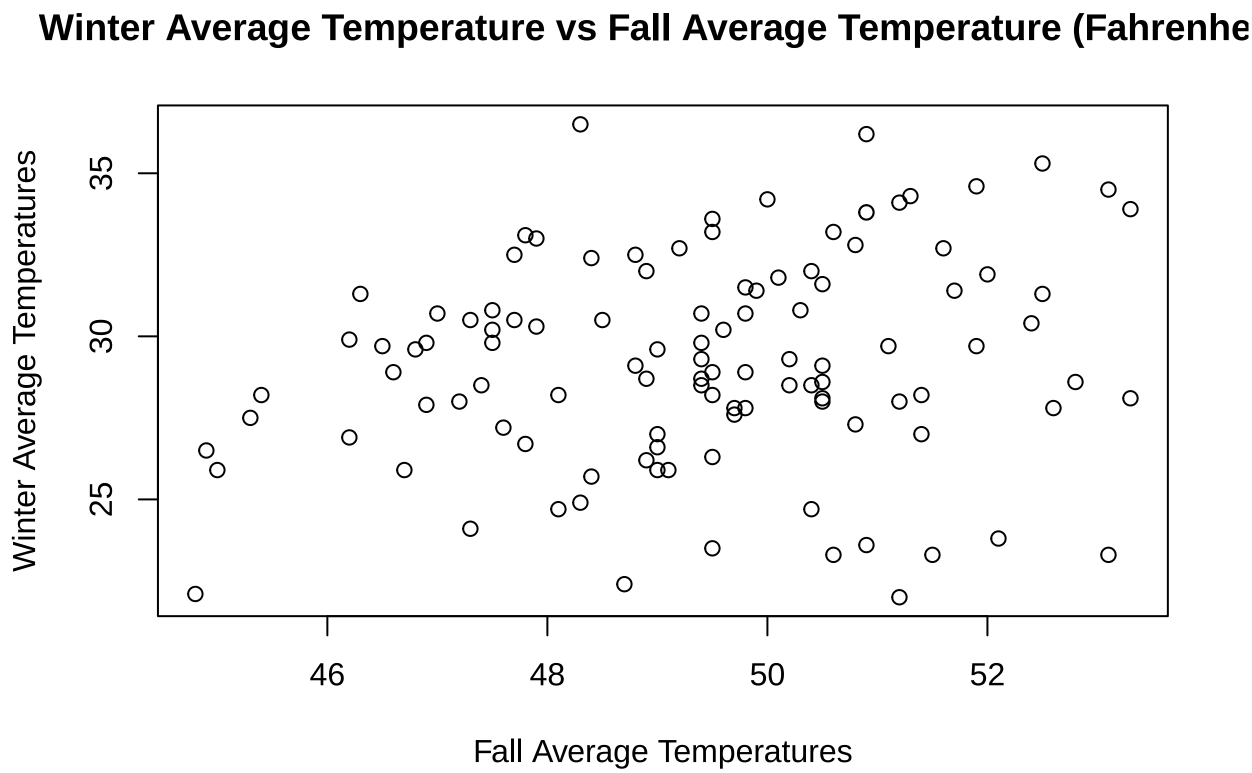
```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(alr4)
```

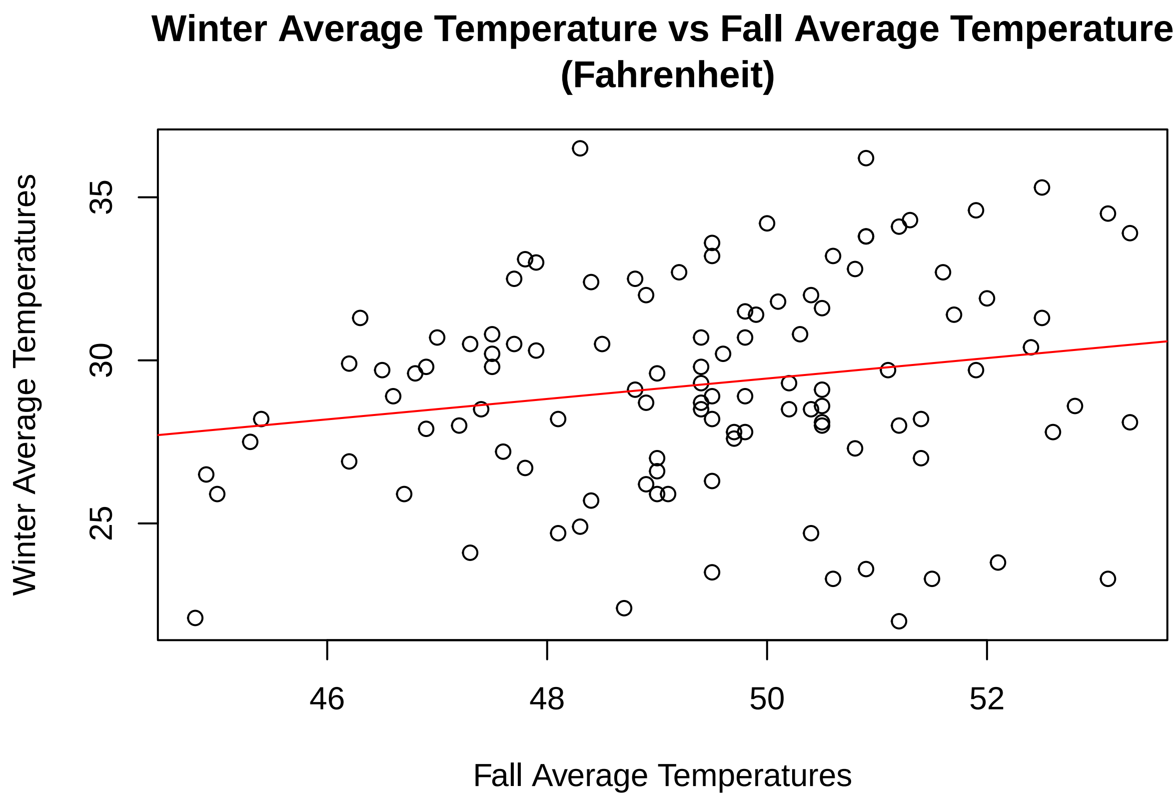
```
## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some

## Loading required package: effects

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
#2.6.1
fcData = ftcollinstemp

plot(fcData$fall, fcData$winter, ylab="Winter Average Temperatures", xlab="Fall Average Temperatures",
     main = "Winter Average Temperature vs Fall Average Temperature (Fahrenheit)")
```




```
# From the plotted data, there seems to be a slight positive correlation for winter vs fall temperature
  # Just from observation, it would be very difficult to determine if the correlation is significant
```

```
#2.6.2

#Creating a linear model and plotting it
fit <- lm(winter~fall, data=fcData)

plot(fcData$fall, fcData$winter, ylab="Winter Average Temperatures", xlab="Fall Average Temperatures",
     main = "Winter Average Temperature vs Fall Average Temperature\n (Fahrenheit)")
abline(fit, col='red')
```




```r
#Hypothesis Testing: NULL Hypothesis: fit slope = 0, alpha = .05
summary(fit)
```

```
##
## Call:
## lm(formula = winter ~ fall, data = fcData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8186 -1.7837 -0.0873  2.1300  7.5896
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.7843     7.5549   1.825   0.0708 .
## fall          0.3132     0.1528   2.049   0.0428 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.179 on 109 degrees of freedom
## Multiple R-squared:  0.0371, Adjusted R-squared:  0.02826
## F-statistic:   4.2 on 1 and 109 DF,  p-value: 0.04284
```

```r
  # Estimate = .3123 | Std. Error = .1528
n <- length(fcData$fall)
t <- (.3123-0)/.1528
```

```r
p <- 2*pt(-t, n-2)
print(p)
```

```
## [1] 0.04337938
```

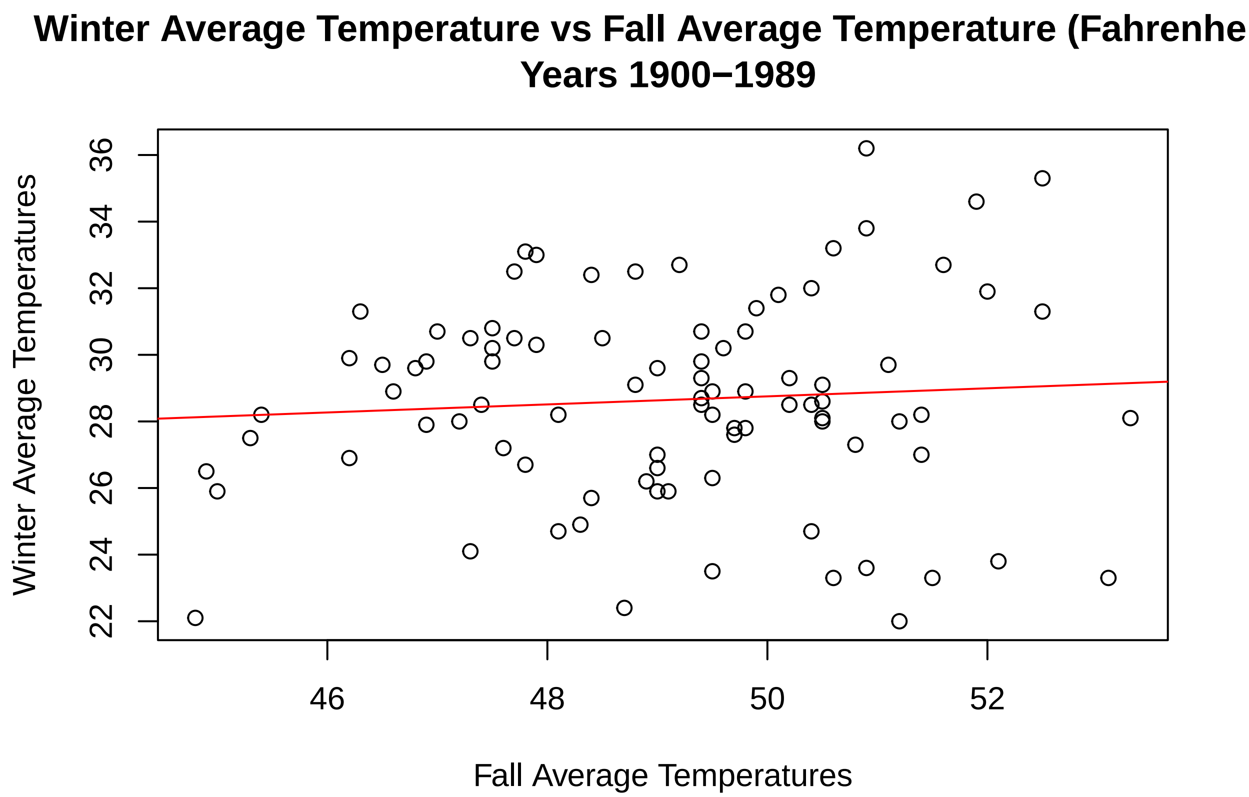
```r
  # We see that p=.0433, therefore we reject the NULL Hypothesis
    # The slope of our regression line is not equal to 0
```

```r
#2.6.3
summary(fit)
```

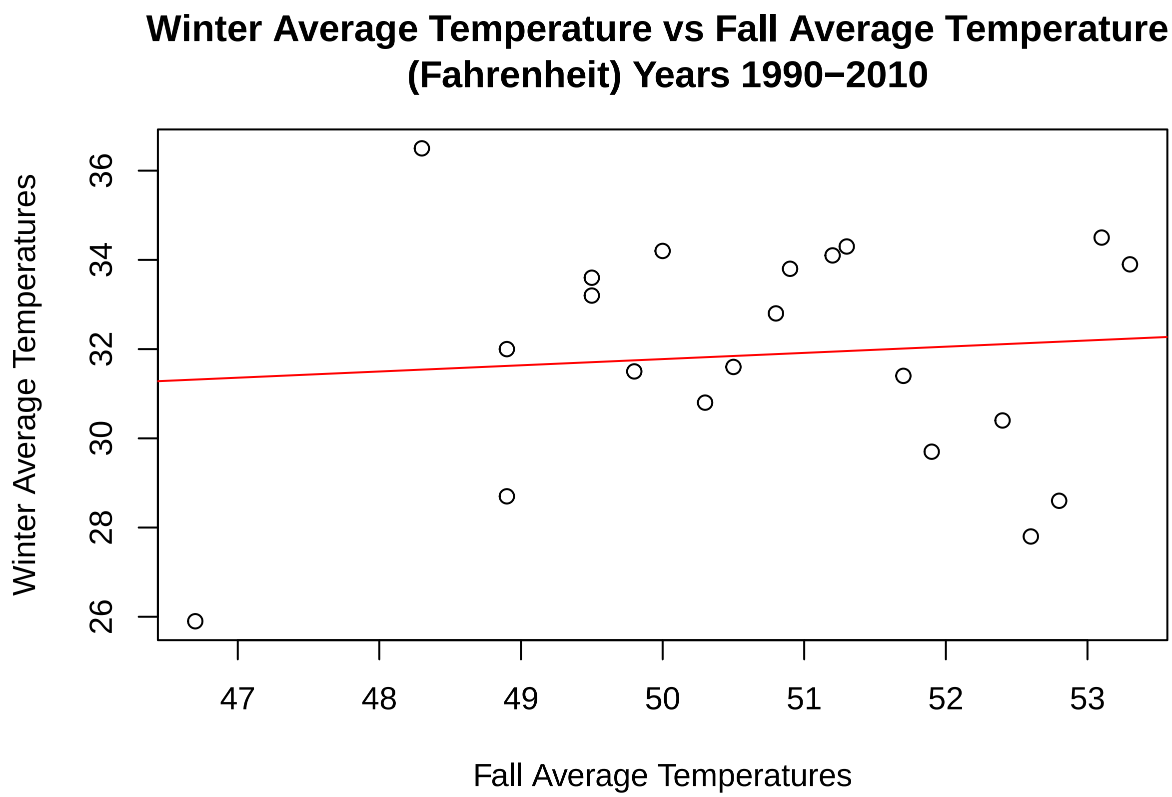
```
##
## Call:
## lm(formula = winter ~ fall, data = fcData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8186 -1.7837 -0.0873  2.1300  7.5896
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.7843     7.5549   1.825   0.0708 .
## fall          0.3132     0.1528   2.049   0.0428 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.179 on 109 degrees of freedom
## Multiple R-squared:  0.0371, Adjusted R-squared:  0.02826
## F-statistic:   4.2 on 1 and 109 DF,  p-value: 0.04284
```

```r
#We can see that the RSS of the data is .0371
#This means there is a 3.71% variance in the data set that is not explained by the regression model
#Generally a low variance means our model is a good fit, with 0% being a perfect fit
  #In our case, we can safely say our model is a good fit
```

```r
#2.6.4
#Creating our two subsets of the data
earlyFcData <- fcData %>%
  filter(year >= 1900 & year <= 1989)

lateFcData <- fcData %>%
  filter(year >= 1990 & year <= 2010)

#Creating simple linear regression models for both of the subsets
fitEarly <- lm(winter ~ fall, data=earlyFcData)
fitLate  <- lm(winter ~ fall, data=lateFcData)

#Creating plots for both of the subsets
plot(earlyFcData$fall, earlyFcData$winter,
     ylab="Winter Average Temperatures", xlab="Fall Average Temperatures",
     main = "Winter Average Temperature vs Fall Average Temperature (Fahrenheit) \n Years 1900-1989")
abline(fitEarly, col='red')
```




```
plot(lateFcData$fall, lateFcData$winter,
     ylab="Winter Average Temperatures", xlab="Fall Average Temperatures",
     main = "Winter Average Temperature vs Fall Average Temperature \n (Fahrenheit) Years 1990-2010")
abline(fitLate, col='red')
```




```r
#Hypothesis testing for our subsets
#NULL Hypothesis: fitEarly slope = 0, alpha = .05
summary(fitEarly)
```

```
##
## Call:
## lm(formula = winter ~ fall, data = earlyFcData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8976 -1.6349  0.0118  2.0079  7.3387
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.7079     8.2600   2.749  0.00725 **
## fall          0.1209     0.1681   0.719  0.47397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.057 on 88 degrees of freedom
## Multiple R-squared:  0.005842,   Adjusted R-squared:  -0.005455
## F-statistic: 0.5171 on 1 and 88 DF,  p-value: 0.474
```

```r
  #Estimate = .1209 | Std. Error = .1681
n <- length(earlyFcData$fall)
```

```r
t <- (.1209-0)/.1681
p1 <- 2*pt(-t, n-2)

#NULL Hypothesis: fitLate slope = 0, alpha = .05
summary(fitLate)
```

```
##
## Call:
## lm(formula = winter ~ fall, data = lateFcData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4174 -1.7097  0.3768  1.8988  4.9602
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.8260    17.7973   1.395    0.179
## fall          0.1390     0.3509   0.396    0.696
##
## Residual standard error: 2.699 on 19 degrees of freedom
## Multiple R-squared:  0.00819,    Adjusted R-squared:  -0.04401
## F-statistic: 0.1569 on 1 and 19 DF,  p-value: 0.6965
```

```r
  #Estimate = .1390 | Std. Error = .3509
n <- length(lateFcData$fall)
t <- (.1390-0)/.3509
p2 <- 2*pt(-t, n-2)

print(p1)
```

```
## [1] 0.4739142
```

```r
print(p2)
```

```
## [1] 0.6964267
```

```r
#In both cases, we fail to reject the NULL Hypothesis, and yes the results of the two different
  #time periods are different from each other (differing std. error, estimate,
  #and variability). Additionally, we can see that the later time period had a
  #steeper slope. But, both subsets failed the hypothesis test
```