

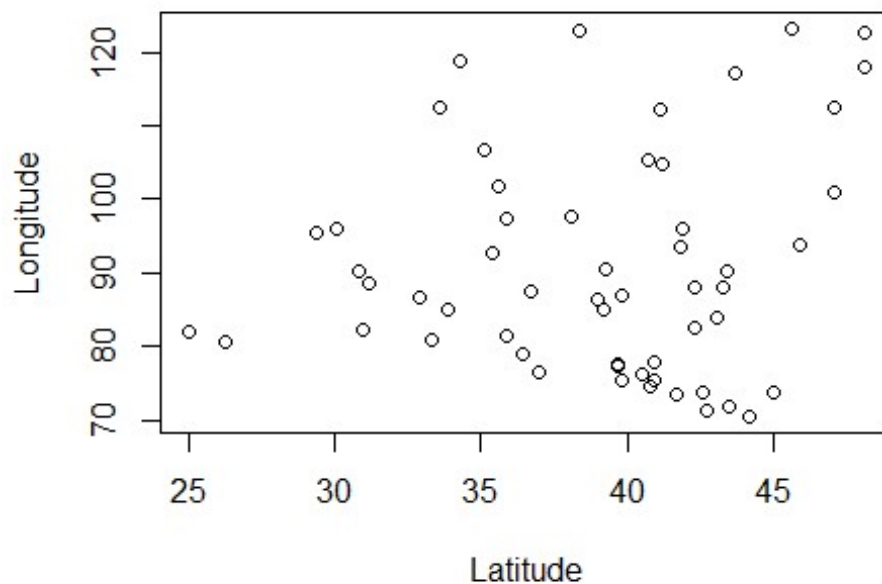
HWK5

Henrique

April 14, 2019

Question 1a)

```
jan = read.csv("january_temp.csv", header = TRUE)
attach(jan)
plot(Lat, Long, xlab= "Latitude", ylab= "Longitude")
```



```
detach(jan)
```

Yes, it looks scattered as I expected.

Question 1b)

```
attach(jan)
estjan <- lm(JanTemp~Lat)
summary(estjan)

##
## Call:
```

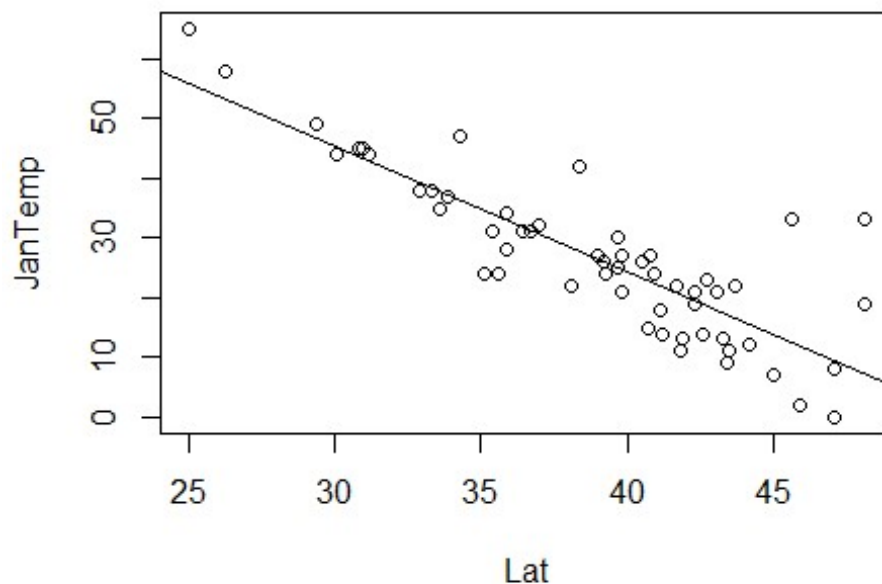
```
## lm(formula = JanTemp ~ Lat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6812  -4.5018  -0.2593   2.2489  25.7434
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 108.7277     7.0561   15.41  <2e-16 ***
## Lat         -2.1096     0.1794  -11.76  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.156 on 54 degrees of freedom
## Multiple R-squared:  0.7192, Adjusted R-squared:  0.714
## F-statistic: 138.3 on 1 and 54 DF,  p-value: < 2.2e-16

plot(JanTemp~Lat)

objects(estjan)

## [1] "assign"      "call"        "coefficients" "df.residual"
## [5] "effects"     "fitted.values" "model"        "qr"
## [9] "rank"        "residuals"   "terms"        "xlevels"

abline(estjan$coefficients)
```



```
JanTemp=108.73-2.11*Lat
```

```
detach(jan)
```

The R squared statistisc is 0.7192.

Question 1c)

```
attach(jan)
```

```
## The following object is masked _by_ .GlobalEnv:
```

```
##
```

```
##      JanTemp
```

```
estjan2 <- lm(JanTemp~Long)
```

```
summary(estjan2)
```

```
##
```

```
## Call:
```

```
## lm(formula = JanTemp ~ Long)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -16.308  -8.325  -2.608   7.878  28.492
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  36.4893      9.4099   3.878 0.000288 ***
```

```
## Long         -0.1098      0.1021  -1.075 0.287085
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 11.33 on 54 degrees of freedom
```

```
## Multiple R-squared:  0.02096,    Adjusted R-squared:  0.002828
```

```
## F-statistic: 1.156 on 1 and 54 DF,  p-value: 0.2871
```

```
plot(JanTemp~Long)
```

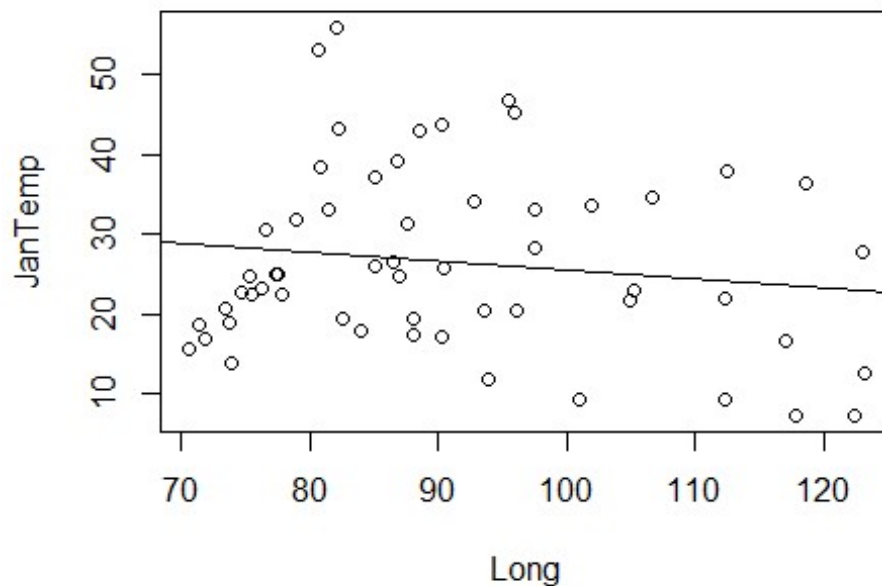
```
objects(estjan2)
```

```
##  [1] "assign"      "call"        "coefficients" "df.residual"
```

```
##  [5] "effects"     "fitted.values" "model"        "qr"
```

```
##  [9] "rank"        "residuals"    "terms"        "xlevels"
```

```
abline(estjan2$coefficients)
```



```
JanTemp2=1.97-2.11*Long
```

```
detach(jan)
```

the R squared statistics is 0.02096.

question 1)d) latitude had a much better prediction, which is shown by its r squared statistic being 1.

Question 1)e)

```
attach(jan)
```

```
## The following object is masked _by_ .GlobalEnv:
```

```
##
```

```
##      JanTemp
```

```
estjan3 <- lm(JanTemp~Long+Lat)
```

```
summary(estjan3)
```

```
## Warning in summary.lm(estjan3): essentially perfect fit: summary may be
## unreliable
```

```
##
```

```
## Call:
```

```
## lm(formula = JanTemp ~ Long + Lat)
```

```
##
```

```
## Residuals:
```

```
##           Min           1Q           Median           3Q           Max
## -1.665e-14 -4.029e-15 -8.970e-16  3.007e-15  4.334e-14
##
## Coefficients:
##           Estimate Std. Error   t value Pr(>|t|)
## (Intercept)  1.087e+02  9.230e-15  1.178e+16  <2e-16 ***
## Long        -1.600e-17  6.999e-17 -2.290e-01    0.82
## Lat         -2.110e+00  1.948e-16 -1.083e+16  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.686e-15 on 53 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 5.995e+31 on 2 and 53 DF, p-value: < 2.2e-16

JanTemp3=108.73-2.11*Lat*Long

detach(jan)
```

Lat was much better at predicting. lat coefficient changed but long didn't

question 1)f)

question 1)g)

```
anova(estjan,estjan2,estjan3)

## Analysis of Variance Table
##
## Model 1: JanTemp ~ Lat
## Model 2: JanTemp ~ Long
## Model 3: JanTemp ~ Long + Lat
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1      54 2765.1
## 2      54 6935.2  0   -4170.1
## 3      53   0.0  1    6935.2 1.1738e+32 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

question 2)

```
x=rnorm(100,10,10)
z = t.test(x)
z$statistic

##           t
## 8.315759

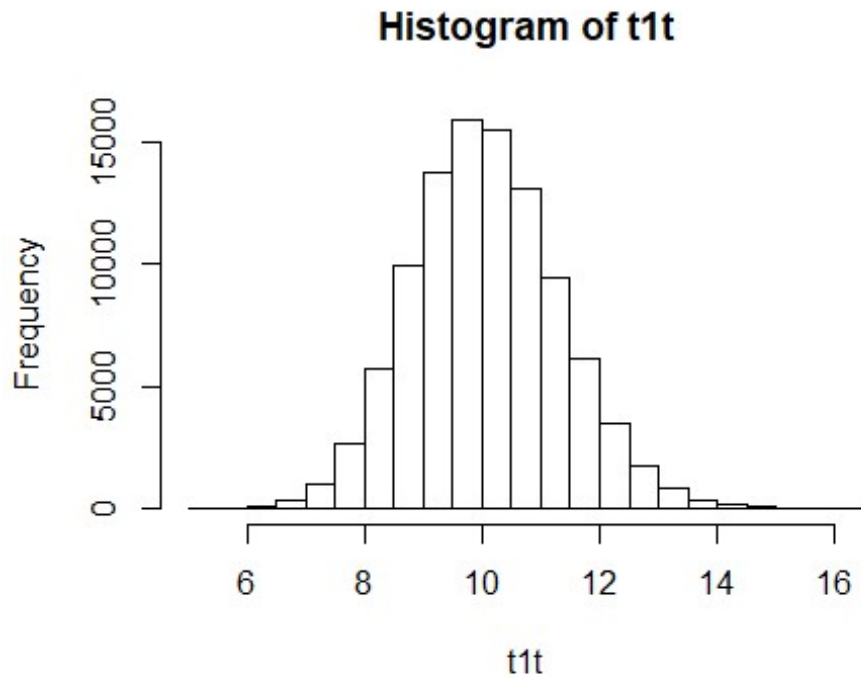
lol <- rep(0,2)
lol[1]<-z$statistic
```

```

t1t <- rep(0,10e4)
for (i in 1:length(t1t)){
  x=rnorm(100,10,10)
  z = t.test(x)
  t1t[i] <- z$statistic
}

hist(t1t)

```



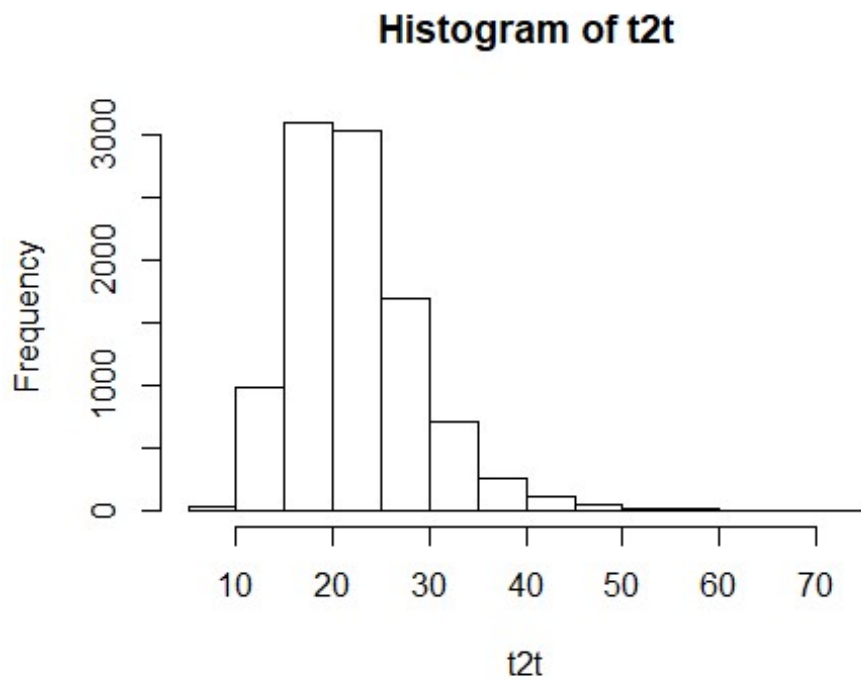
question 3a)

```

t2t <- rep(0,10000)
for (i in 1:10000){
  x<-(rnorm(100,5,1)+rnorm(10,5,5))
  z = t.test(x)
  t2t[i] <- (z$statistic)
}

hist(t2t)

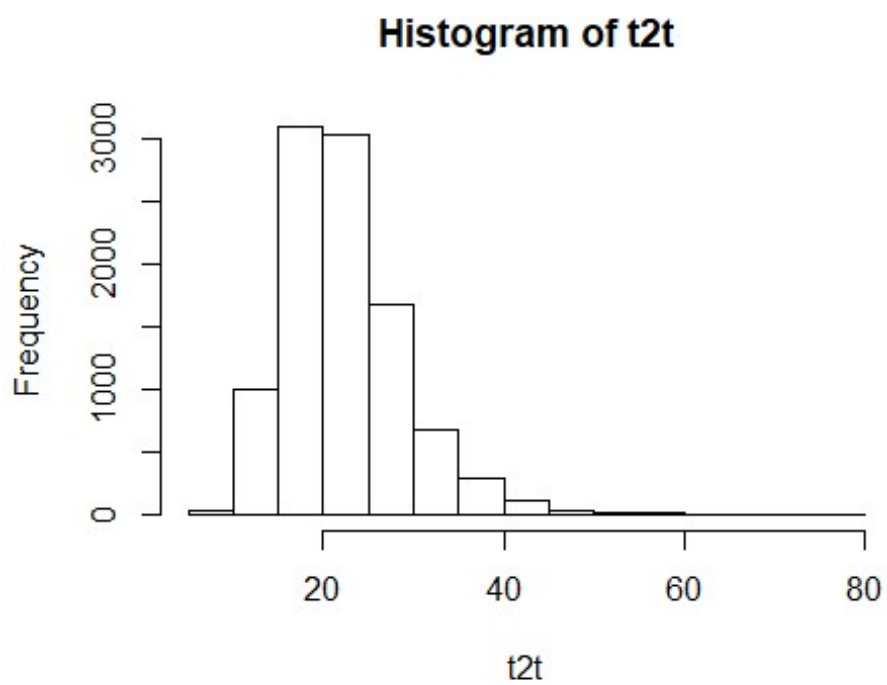
```



3b)

```
t2t <- rep(0,10000)
for (i in 1:10000){
  x<-(rnorm(100,5,1)+rnorm(10,5,5))
  z = t.test(x,var.equal = TRUE)
  t2t[i] <- (z$statistic)
}

hist(t2t)
```



part c looks less skewed to the right when compared to part b