Inan Khan
Assignment 5

1.

```
> t.test(r_sample$Height,r_sample2$Height)

        Welch Two Sample t-test

data:  r_sample$Height and r_sample2$Height
t = 2.9979, df = 36.469, p-value = 0.004872
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.685103 3.546662
sample estimates:
mean of x mean of y
 70.05127  67.93538
```

**First I made two separate tables, r_sample is 20 samples of male data from weight.height and r_sample2 is 20 samples of female data. Then I did the t-test and the results are shown

2.
After I added 0.1 once

```
        Welch Two Sample t-test

data:  grouped_data$group1 and grouped_data$group2
t = 2.8562, df = 36.469, p-value = 0.007037
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.585103 3.446662
sample estimates:
mean of x mean of y
 70.05127  68.03538
```

After I added 0.1 twice

```
        Welch Two Sample t-test

data:  grouped_data$group1 and grouped_data$group2
t = 2.7145, df = 36.469, p-value = 0.01008
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.485103 3.346662
sample estimates:
mean of x mean of y
 70.05127  68.13538
```

As we can see I got it below 0.5 after adding 0.1 twice to all the female values

3. If I made every female value equal to the mean, then It would still be the same on how much I have to add to show lack of significance

4.Obviously when you add to the female group it makes it closer to the male group so the differences in average would go down, but the standard deviation would still be the same.

A. Top 20 and bottom 20 males grouped by Height

```
> top_male <- male_wh[with(male_wh,order(-Height)),]
> top_male <- top_male[1:10,]
> top_male <- male_wh[with(male_wh,order(-Height)),]
> top_male <- top_male[1:20,]
> bottom_male <- male_wh[with(male_wh,order(Height)),]
> bottom_male <- bottom_male[1:20,]
> |
```

Top 20 and bottom 20 females

```
> top_female <- female_wh[with(female_wh,order(-Height)),]
> top_female <- top_female[1:20,]
> bottom_female <- female_wh[with(female_wh,order(Height)),]
> bottom_female <- bottom_female[1:20,]
```

Top male and female pairwise t-test

```
> pairwise.t.test(top_male$Height,top_female$Height)

        Pairwise comparisons using t tests with pooled SD

data:  top_male$Height and top_female$Height

                  76.6909627296268
76.7098348578592 -
76.7199849250776 -
76.732446458109  -
76.8063436350608 -
76.8426788279609 -
76.8487612691053 -
76.8668549581823 -
77.0083360356505 -
77.0673550156967 -
77.1008721027022 -
77.1608008945486 -
77.4466199509585 -
77.4655691046729 -
```

Bottom male and female pairwise t-test

```
> pairwise.t.test(bottom_male$Height,bottom_female$Height)

        Pairwise comparisons using t tests with pooled SD

data:  bottom_male$Height and bottom_female$Height

                  58.4069049317498 59.3806495599258 59.8680779342708 59.9386496524854 59.9818650708608 60.217018557323
59.3806495599258 -                -                -                -                -                -
59.8680779342708 -                -                -                -                -                -
59.9386496524854 -                -                -                -                -                -
59.9818650708608 -                -                -                -                -                -
60.217018557323  -                -                -                -                -                -
60.2437181028691 -                -                -                -                -                -
60.267190118023  -                -                -                -                -                -
60.2691078279205 -                -                -                -                -                -
```

1. From the results of A and B i can figure that the range of male values between minimum and maximum is much larger than for females.

2. There would probably have to be several hundred males to delete before you can make the ranges of heights between the two groups similar.
3. As for females, you would likely have to add values on either side of the spectrum to make the range in values more similar to males. Doing this would consequently also make the averages more similar.
4. It would be better to use the sets without deletions to prove that males are taller than females because the full set emphasizes the differences between the two groups such as the difference in average and range in heights.