206\_Assignment\_5

Tess Hooper

12/2/2018

## Step 1. Load packages and read in data.

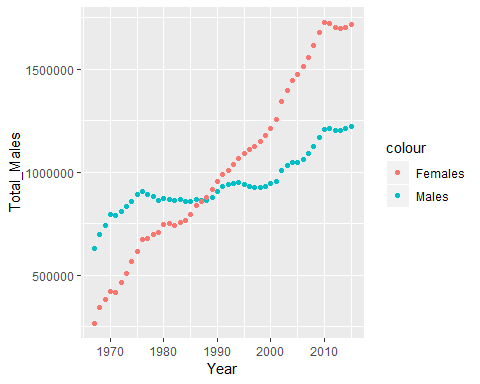
Part 1: Male and female graduate enrollment (1967 - 2015)

## Compare trends in total graduate enrollment for males and females (use Totals) in the United States from 1967 - 2015. Describe your results statistically, graphically and in text.

1. Exploratory Data Analysis

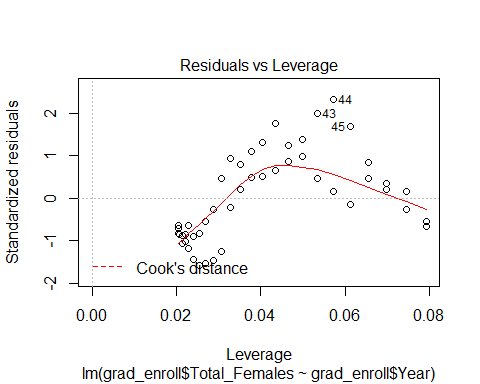
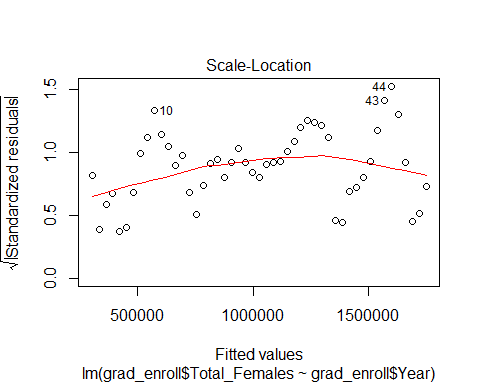
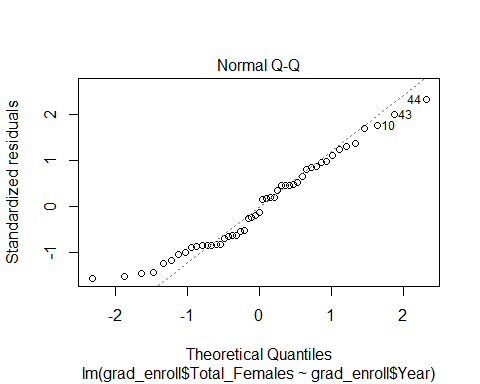
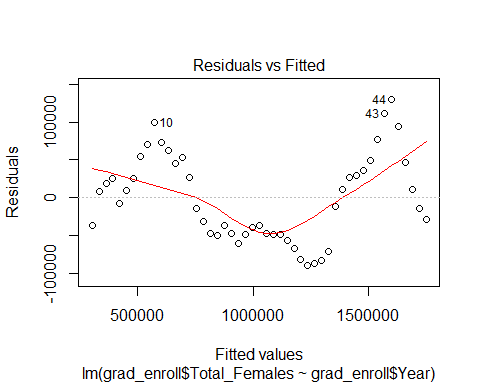
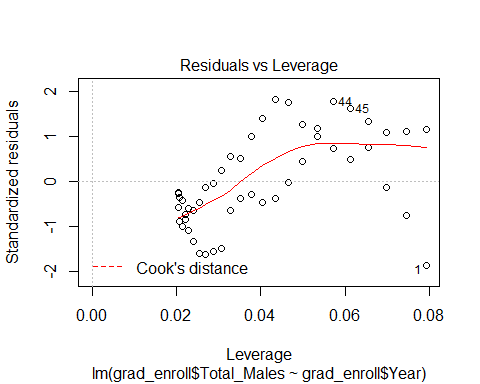
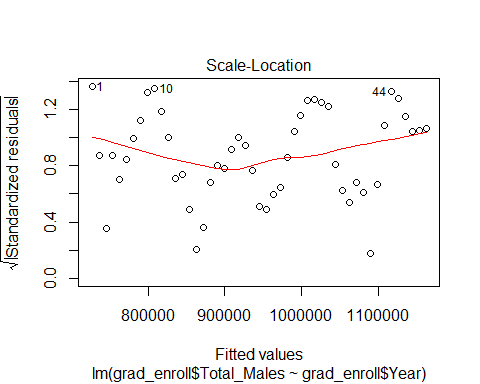
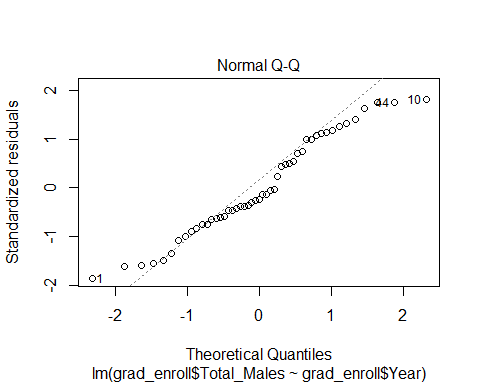
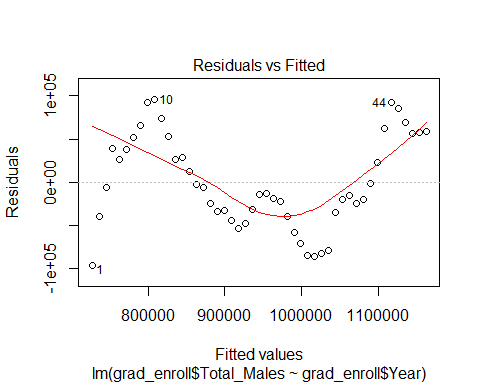
* Summary data table with graduate enrollment for males and females, part-time/full-time, private/public universites
* Scatter plots for full-time/part-time and gender, private/public universities and gender
* Regression line for males and females (we need to do a regression because we need the line of best fit to find residuals in order to make sure our model fits our assumptions - residuals are any point that isn’t on the line of best fit )

1. Question: How does enrollment in regards to gender change within these categories over time?



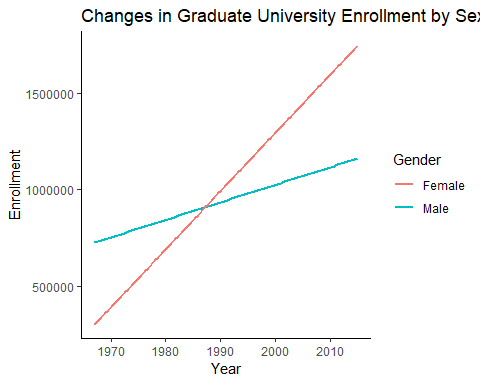
##   
## Call:  
## lm(formula = grad\_enroll$Total\_Females ~ grad\_enroll$Year)  
##   
## Coefficients:  
## (Intercept) grad\_enroll$Year   
## -58955502 30126

##   
## Call:  
## lm(formula = grad\_enroll$Total\_Males ~ grad\_enroll$Year)  
##   
## Coefficients:  
## (Intercept) grad\_enroll$Year   
## -17112153 9069



##   
## Call:  
## lm(formula = grad\_enroll$Total\_Females ~ grad\_enroll$Year)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -89397 -48101 -7633 45267 129727   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.896e+07 1.161e+06 -50.77 <2e-16 \*\*\*  
## grad\_enroll$Year 3.013e+04 5.832e+02 51.66 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 57730 on 47 degrees of freedom  
## Multiple R-squared: 0.9827, Adjusted R-squared: 0.9823   
## F-statistic: 2669 on 1 and 47 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = grad\_enroll$Total\_Males ~ grad\_enroll$Year)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -96461 -34861 -12841 51876 95766   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -17112153 1087024 -15.74 <2e-16 \*\*\*  
## grad\_enroll$Year 9069 546 16.61 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 54050 on 47 degrees of freedom  
## Multiple R-squared: 0.8545, Adjusted R-squared: 0.8514   
## F-statistic: 276 on 1 and 47 DF, p-value: < 2.2e-16



<<<<<<< HEAD

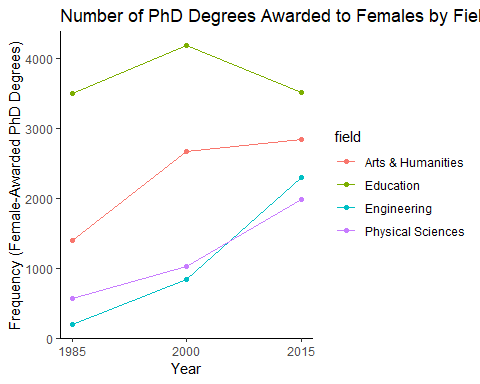
# ##Part 2. Shifts in female PhD recipients by field (1985, 2000, and 2015).

Part 2. Shifts in female PhD recipients by field (1985, 2000, and 2015). >>>>>>> b400d8b7001a71f3d0af765920a26a74c22bc5a0

1. Describe if and how there was a shift in PhDs awarded to females in Physical and Earth Sciences, Engineering, Education, and Humanities & Arts in 1985, 2000, and 2015.
2. Describe in a graph or table.
3. Describe in text and statistically.

# Note: several ways to interpret this question: decide which you think is/are most interesting. state which you are asking/answering in your report.

## # A tibble: 12 x 5  
## field sex year number percent  
## <chr> <chr> <int> <dbl> <dbl>  
## 1 Physical Sciences female 1985 569 0.163  
## 2 Engineering female 1985 198 0.063  
## 3 Education female 1985 3491 0.518  
## 4 Arts & Humanities female 1985 1392 0.409  
## 5 Physical Sciences female 2000 1022 0.252  
## 6 Engineering female 2000 838 0.158  
## 7 Education female 2000 4179 0.649  
## 8 Arts & Humanities female 2000 2672 0.49   
## 9 Physical Sciences female 2015 1988 0.336  
## 10 Engineering female 2015 2301 0.232  
## 11 Education female 2015 3502 0.685  
## 12 Arts & Humanities female 2015 2832 0.506



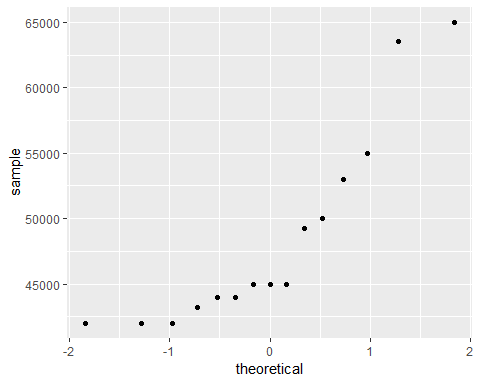
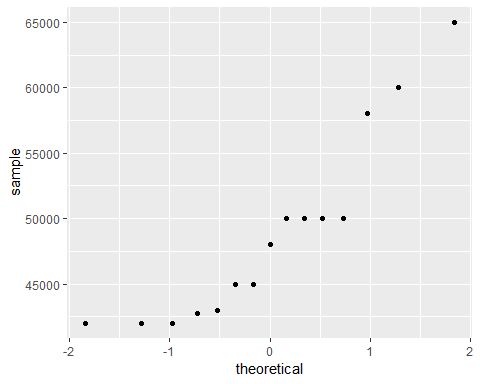
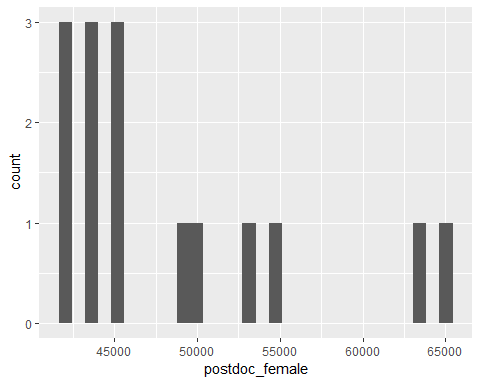
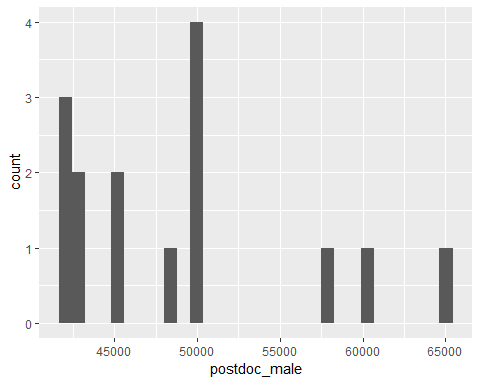
#1. Use chi-squared to explore changes in PhD proportions between fields over time.  
  
  
#Testing if our values are significantly different from proportions we would expect in a normally-distributed population.  
  
#Null hypothesis (H0): There is no significant difference between the observed and the expected value.  
#Alternative hypothesis (Ha): There is a significant difference between the observed and the expected value  
  
  
  
#interested in showing the changes in proportions of number of degrees between the categories across the 3 years. Also should show total increase or decrease in number of degrees per category.

## Part 3: Male and female salaries for starting postdoctoral and other employment positions (2015).

Compare median salaries for male and female doctorate recipients in 2015. Answer these two questions:

A. Does median salary differ significantly between male and female starting postdoc positions?

B. Does median salary differ significantly between male and female PhD recipients in non-postdoc employment positions?



## Part 4. Exploring academic salaries for professors in U.S. colleges.

Develop a model describing faculty salary based on data for faculty sex, rank, years in current position, field, and number of years since doctoral degree was earned.