Lab Assignment 4

Nora Quick

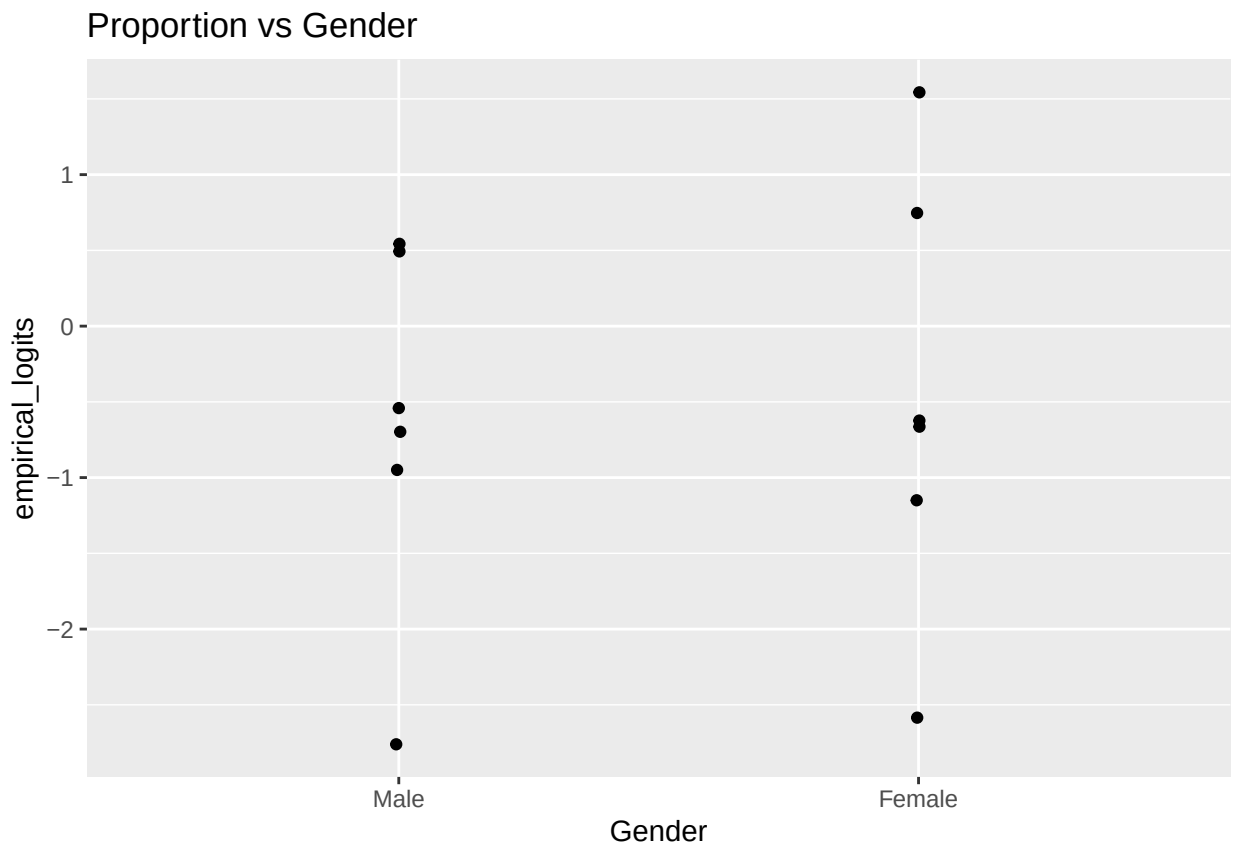
Questions:

```
library(Sleuth3)
library(tidyverse)
library(ggplot2)
library(arm)
library(vcdExtra)
library(magrittr)
```

(a) Construct an informative ggplot() of the empirical logits of admission proportion vs gender and department. It's up to you what aesthetics to map to which variables – there is more than one right answer here.

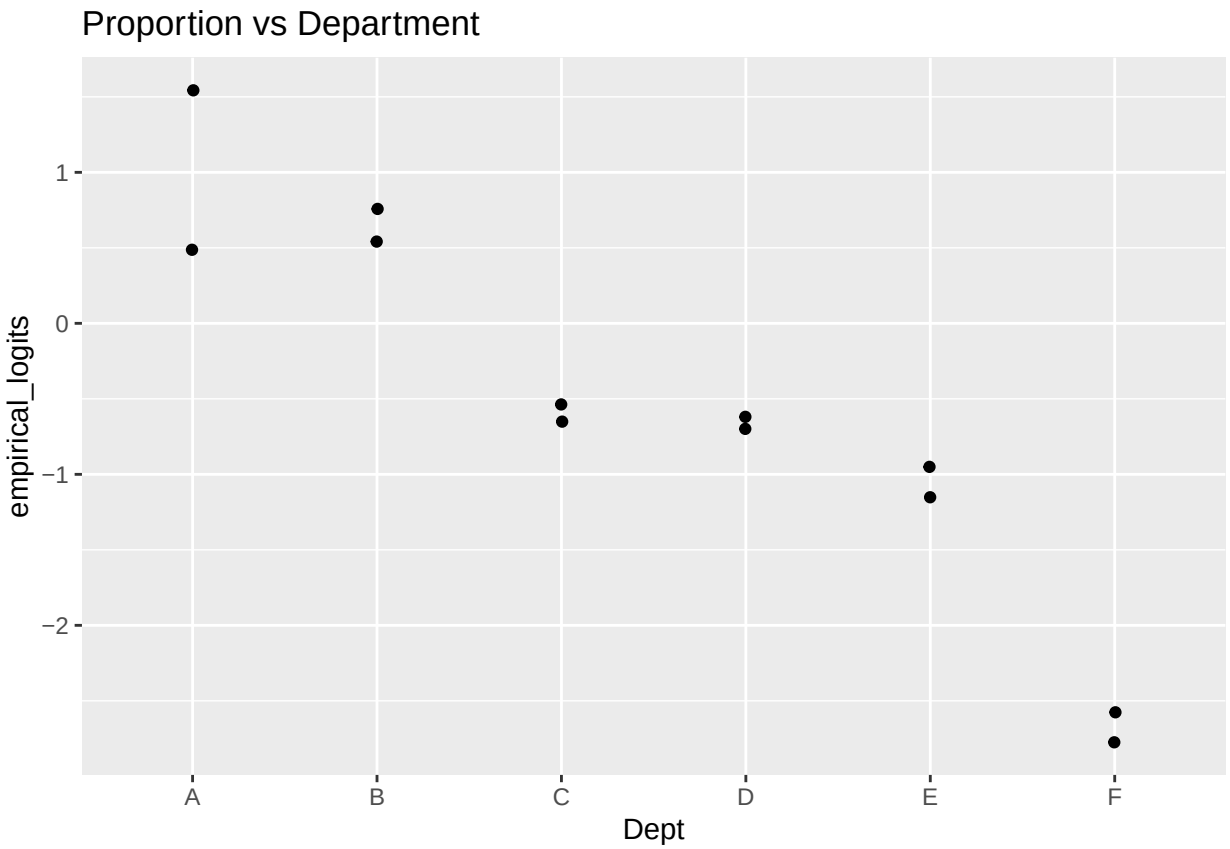
```
dat = data.frame(UCBAdmissions)
dat2 = spread(dat, Admit, Freq)

dat2$Total = dat2$Admitted + dat2$Rejected

empirical_logits <- with(dat2, qlogis(Admitted/Total))

ggplot(data = dat2, aes(x = Gender, y = empirical_logits))+
  geom_jitter(width = 0.005, height = 0.01) +
  geom_smooth(se = FALSE) +
  ggtitle("Proportion vs Gender")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data = dat2, aes(x = Dept, y = empirical_logits))+
  geom_jitter(width = 0.005, height = 0.01) +
  geom_smooth(se = FALSE) +
  ggtitle("Proportion vs Department")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



(b) Based on your plot from (a), which variable (gender or department) appears to account for more of the variability in admissions? Explain.

Based on the above graphs it appears that department has more variability in admissions. Between the two graphs we see much more of a difference between the different x-axis variables. The genders largly overlap with just a few outliers.

(c) Fit an appropriate (binomial) logistic regression model for admissions. What is the estimated dispersion parameter? Is there evidence of lack of fit?

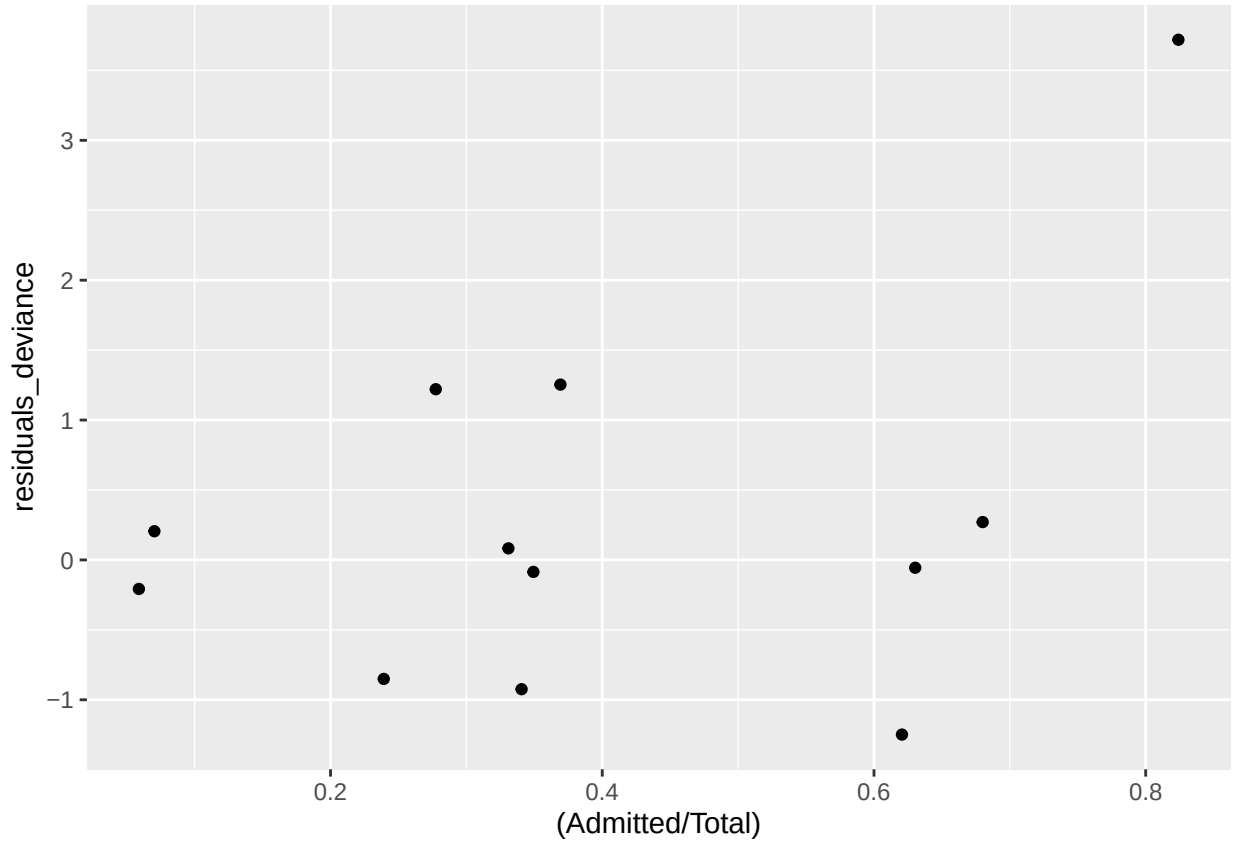
```
mod <- glm(data = dat2, (Admitted/Total) ~ Gender + Dept, weights = Total, family = "binomial")
LRstats(mod)
```

```
## Likelihood summary table:
##        AIC    BIC LR Chisq Df Pr(>Chisq)
## mod 103.14 106.54   20.204  5   0.001144 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the lab we can look at chi-squared for the estimated dispersion parameter which is 20.2. There is some evidence of a lack of fit.

(d) Construct a plot of residuals vs fitted values (try just `plot`-ing your fitted model object). From this plot, can you identify a source for any fit problems encountered in part (c)?

```
dat2$residuals_deviance <- residuals(mod)

ggplot(data = dat2, aes((Admitted/Total),residuals_deviance)) + geom_point()
```



There is no obvious pattern for this model.

(d) Refit the binomial model above, but excluding the data from department A. Now what is the estimated dispersion parameter? Based on the p-value, what would you conclude about the effect of Gender on admissions (to departments other than A) using this model?

```
dat3 <- subset(dat2, Dept != "A")

mod2 <- glm(data = dat3, (Admitted/Total) ~ Gender + Dept, weights = Total, family = "binomial")
LRstats(mod2)
```

```
## Likelihood summary table:
##         AIC    BIC LR Chisq Df Pr(>Chisq)
## mod2 71.791 73.606   2.5564  4     0.6346
```

The estimated dispersion parameter is 2.6 and based on the p-value this is a much better fit. In addition to that I would say that the p-value indicates that there is some effect from gender on the admissions rate.

(e) Using the quasibinomial model for all departments, what do you conclude about the effect of Gender on admissions?

```
mod3 <- glm(data = dat2, (Admitted/Total) ~ Gender + Dept, weights = Total, family = "quasibinomial")
LRstats(mod2)
```

```
## Likelihood summary table:
##         AIC    BIC LR Chisq Df Pr(>Chisq)
## mod2 71.791 73.606   2.5564  4     0.6346
```

```
confint(mod3)
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %     97.5 %
## (Intercept)   0.3225043  0.8480077
## GenderFemale -0.2061476  0.4093268
## DeptB        -0.4600039  0.3764602
## DeptC        -1.6723809 -0.8606612
## DeptD        -1.7015531 -0.8959751
## DeptE        -2.2280004 -1.2674256
## DeptF        -3.9984153 -2.6961352
```

I conclude that both variables play a role in the variance of being admitted, however, department has a larger role.