Lab Assignment 3

Nora Quick

```
library(vcdExtra)
library(tidyverse)
library(magrittr)
```

1. Install and load the *AER* package, and read the help file for the `ResumeNames` data.

```
library(AER)
```

```
?ResumeNames

data("ResumeNames")
#summary(ResumeNames)
```

2. Come up with a question about the probability of `callback` (the binary response) that can be answered using at least one (but no more than 5) of the 26 available predictor variables.

Does callback probability differ between male and female applicants by industry or position?

3. Use a logisitic regression model to address the question you posed in 2. Be sure to examine the fit of your model, and write a few sentences about your interpretation of the model as it addresses the question you posed.

```
gender_callback <- glm(call ~ gender + industry + wanted, family = binomial(link = "logit"), data = Resu
summary(gender_callback)
```

```
##
## Call:
## glm(formula = call ~ gender + industry + wanted, family = binomial(link = "logit"),
##     data = ResumeNames)
##
```

```
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6499  -0.4182  -0.3935  -0.3605   2.4852
##
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -2.998220   0.273695 -10.955  < 2e-16
## genderfemale                            -0.028511   0.152641  -0.187  0.85183
## industrytransport/communication          1.023606   0.328726   3.114  0.00185
## industryfinance/insurance/real estate    0.271677   0.293404   0.926  0.35447
## industrytrade                            0.296977   0.253439   1.172  0.24128
## industrybusiness/personal services       0.436614   0.242498   1.800  0.07178
## industryhealth/education/social services 0.648648   0.255021   2.544  0.01097
## industryunknown                          0.354930   0.257232   1.380  0.16765
## wantedsupervisor                        -0.014769   0.260176  -0.057  0.95473
## wantedsecretary                          0.182287   0.186979   0.975  0.32961
## wantedoffice support                     0.555636   0.208046   2.671  0.00757
## wantedretail sales                       0.209946   0.197168   1.065  0.28696
## wantedother                             -0.006485   0.210637  -0.031  0.97544
##
## (Intercept)                             ***
## genderfemale
## industrytransport/communication          **
## industryfinance/insurance/real estate
## industrytrade
## industrybusiness/personal services       .
## industryhealth/education/social services *
## industryunknown
## wantedsupervisor
## wantedsecretary
## wantedoffice support                     **
## wantedretail sales
## wantedother
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2726.9  on 4869  degrees of freedom
## Residual deviance: 2698.9  on 4857  degrees of freedom
## AIC: 2724.9
##
## Number of Fisher Scoring iterations: 5
```

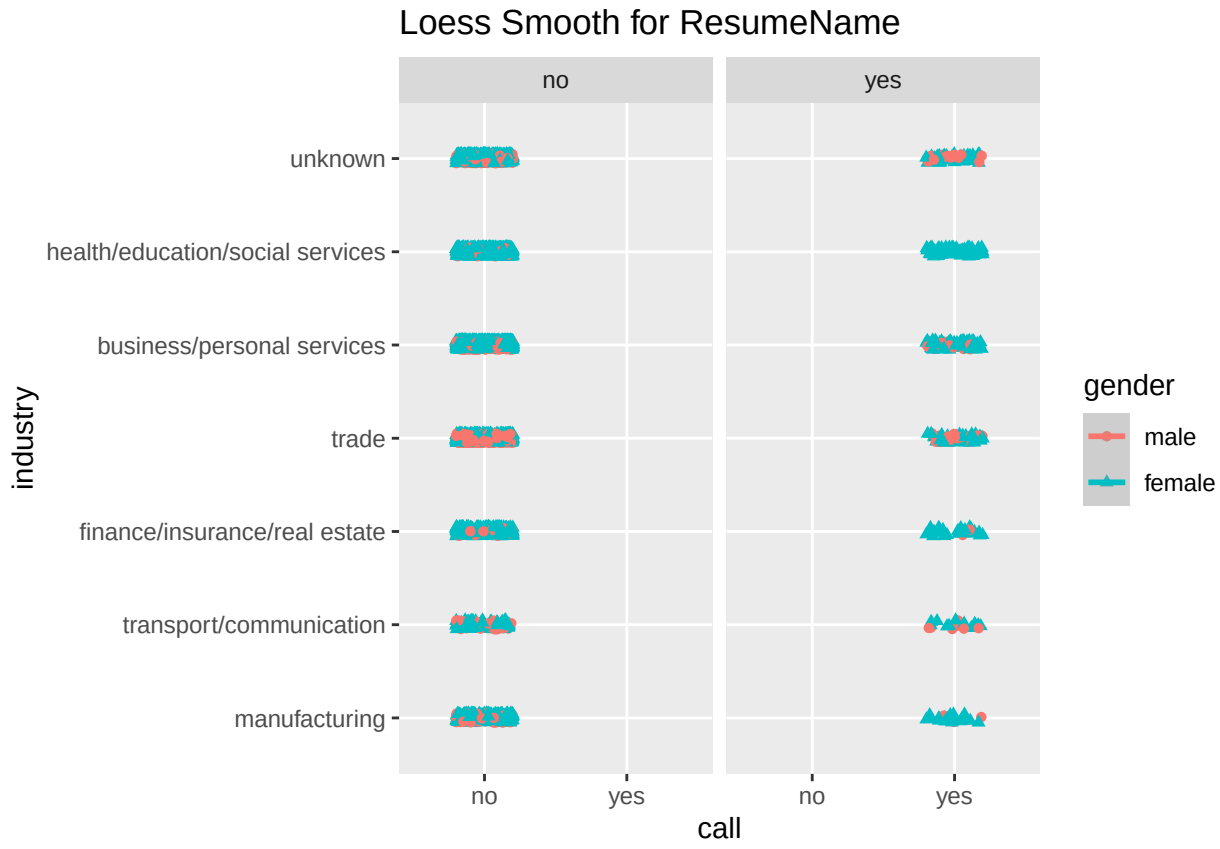
```
ggplot(data = ResumeNames) +
  # plot jittered data
  geom_jitter(aes(x = call,
                  y = industry,
                  color = gender,
                  shape = gender),
                  height = 0.05, width = 0.2) +

  # plot loess smoother
  geom_smooth(aes(x = call,
```

```
                y = industry,
                 color = call)) +

facet_grid(.~call) +
  ggtitle("Loess Smooth for ResumeName")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

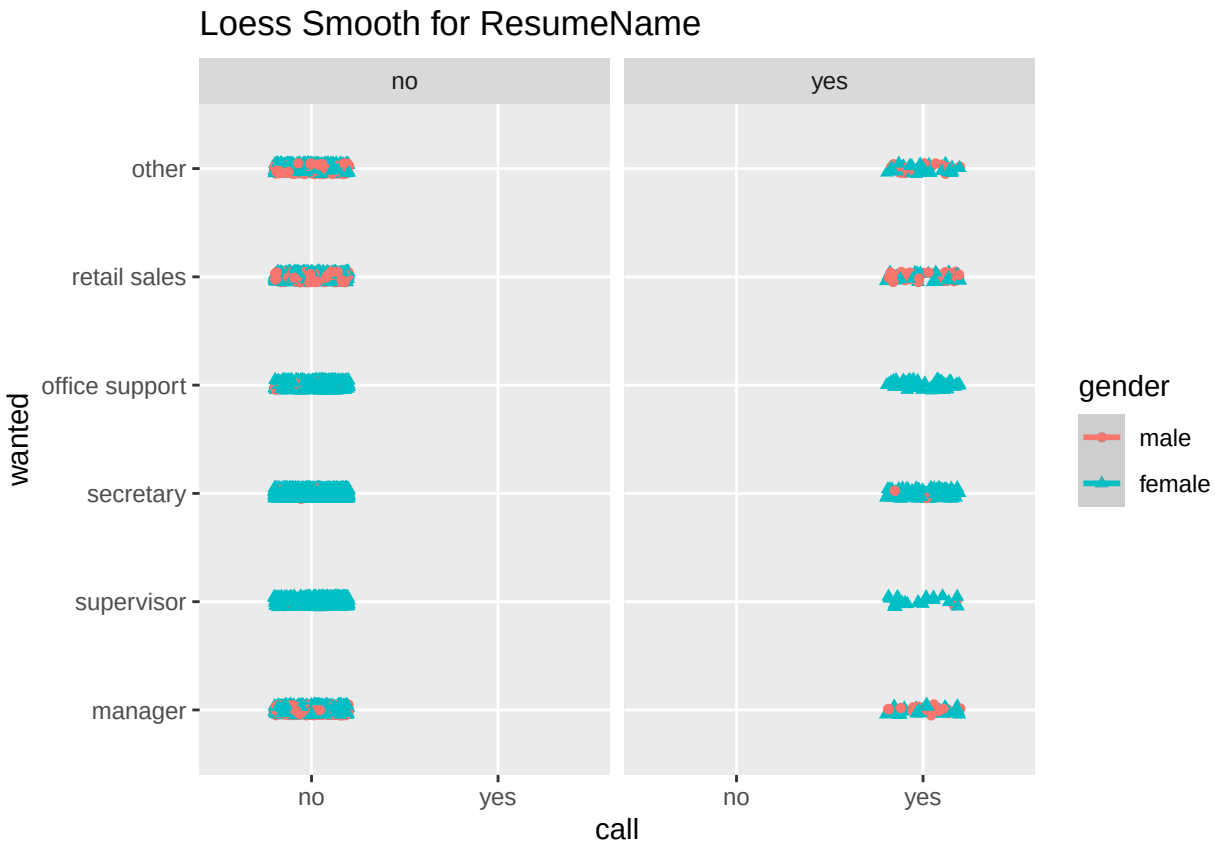



```
ggplot(data = ResumeNames) +
  # plot jittered data
  geom_jitter(aes(x = call,
                  y = wanted,
                  color = gender,
                  shape = gender),
                  height = 0.05, width = 0.2) +

  # plot loess smoother
  geom_smooth(aes(x = call,
                y = wanted,
                 color = call)) +

facet_grid(.~call) +
  ggtitle("Loess Smooth for ResumeName")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



As we can see by both the glm output and both graphs there is a probability difference for callbacks in both industry and position. Our logistic model shows significance in industrytransport/communication, industryhealth/education/social services, and wantedoffice support. As we can see these correlate to the most female dominant callback industires and positions. These are shown to be more probable positions for woman than men and we could also determine this based on the society we live in. We largely see women in health care positions, educational positions, and secretarial positions due to historical reasons.