241 Final Project Analysis

Samir Datta

April 14, 2018

1. Load data

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.3.2
```

```
library(ggplot2)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```
library(multiwayvcov)
library(sandwich)

dodge = position_dodge(width=0.9)
theme_set(theme_gray(base_size = 13))

#setwd('C:/Users/Samir/Documents/MIDS/ExperimentsCausalityW18')
```

```
source('clean_data.R')
```

```
## Warning: package 'data.table' was built under R version 3.3.2
```

```
## Warning in ifelse(!is.na(as.numeric(alldata$date_applied)), 1, 0): NAs
## introduced by coercion
```

```
print(paste('total # of applicatons sent:', nrow(alldata)))
```

```
## [1] "total # of applicatons sent: 300"
```

```
print(paste('total # of usable applicatons (where both candidates succesfully submitted):', nrow(alldata
```

```
## [1] "total # of usable applicatons (where both candidates succesfully submitted): 210"
```

```
print(paste('total # of usable west coast applicatons:', nrow(alldata[alldata$both_applications_valid==
                                                                      alldata$coast=='West',])))
```

```
## [1] "total # of usable west coast applicatons: 96"
```

```
print(paste('total # of usable east coast applicatons:', nrow(alldata[alldata$both_applications_valid==
                                                                      alldata$coast=='East',])))
```

```
## [1] "total # of usable east coast applicatons: 114"
```

Covariate checks

```
print(table(alldata[alldata$both_applications_valid==1,]$size_bin,
          alldata[alldata$both_applications_valid==1,]$coast))
```

```
##
##          East West
##   Large    36   46
##   Medium   44   34
##   Small    34   16
```

```
  chisq.test(alldata[alldata$both_applications_valid==1,]$size_bin,
          alldata[alldata$both_applications_valid==1,]$coast)
```

```
##
##  Pearson's Chi-squared test
##
## data:  alldata[alldata$both_applications_valid == 1, ]$size_bin and alldata[alldata$both_applications
## X-squared = 7.4938, df = 2, p-value = 0.02359
```

We are no longer balanced in terms of the distribution of company size by coast. We found more large west coast companies, and more medium/small east coast companies.

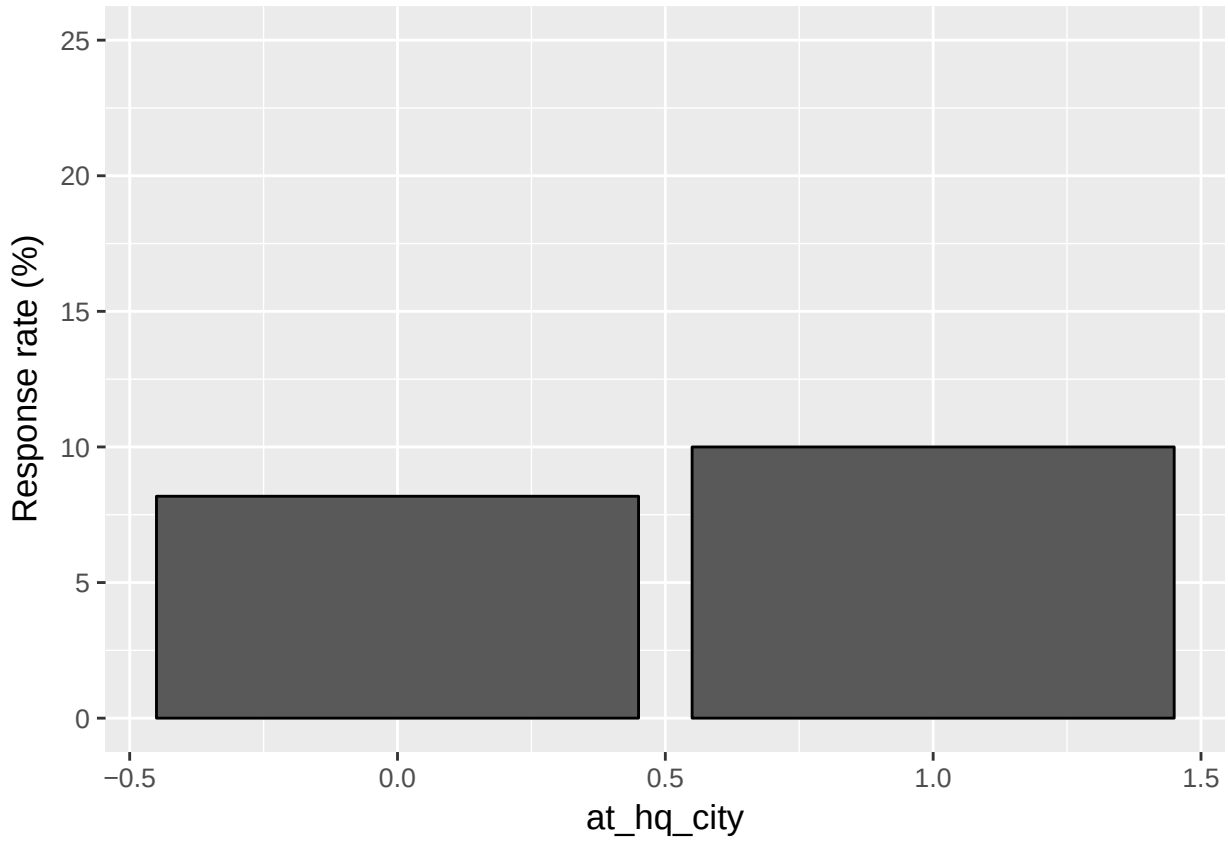
Response rate bar graph - skeleton

```
alldata_agg <- with(alldata[alldata$both_applications_valid==1,],
                    aggregate(job_id,
                              list(call_back_binary=call_back_binary,
                                   at_hq_city=at_hq_city), length))

alldata_agg$response_rate <- NA
alldata_agg[alldata_agg$call_back_binary==1,]$response_rate <-
  100*alldata_agg[alldata_agg$call_back_binary==1,]$x/
  (alldata_agg[alldata_agg$call_back_binary==1,]$x+
     alldata_agg[alldata_agg$call_back_binary==0,]$x)


ggp <- ggplot(alldata_agg[!is.na(alldata_agg$response_rate)], aes(x=at_hq_city, y=response_rate))
ggp + geom_bar(stat="identity", color="black", position=dodge)+
  ylab('Response rate (%)')+ylim(c(0,25))
```

```
## Warning: Removed 2 rows containing missing values (geom_bar).
```

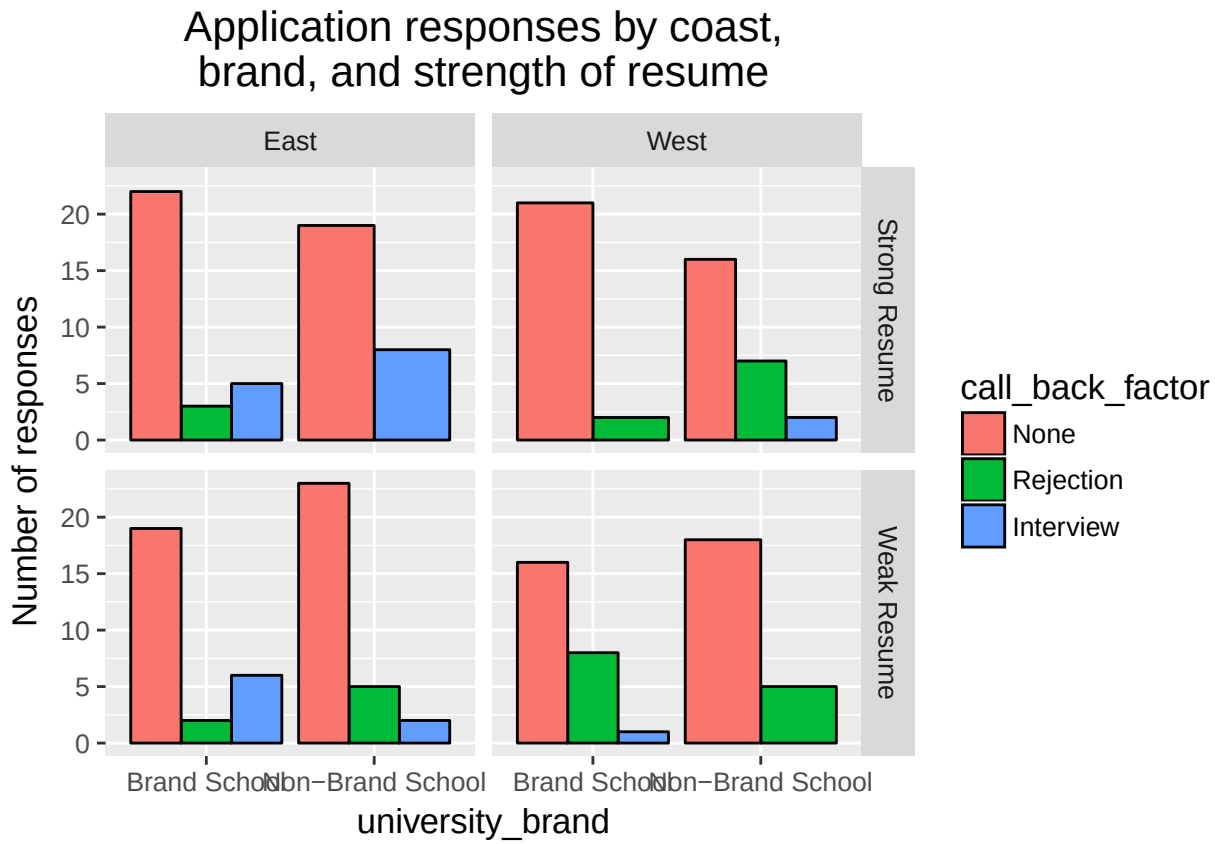



```
  ggtitle('Application response rates')+
  theme_update(plot.title = element_text(hjust = 0.5))
```

```
## NULL
```

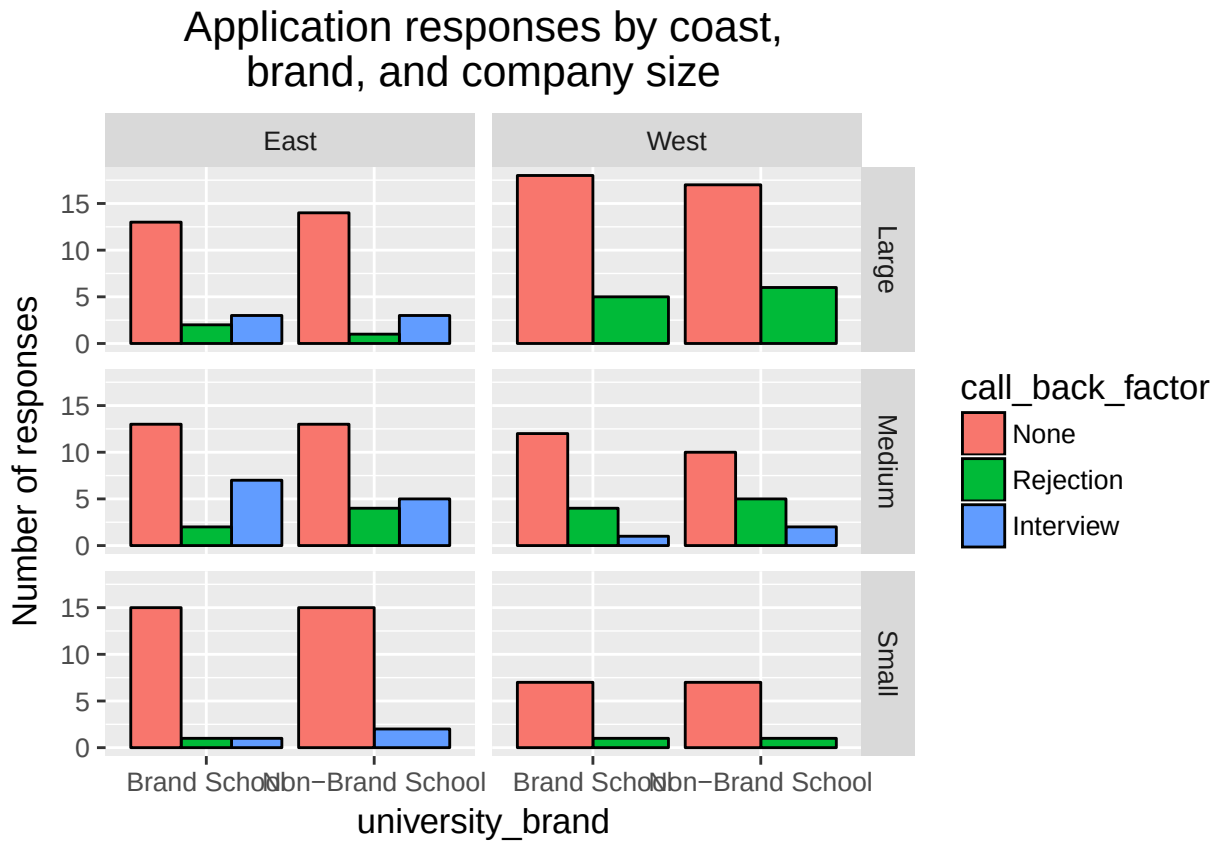
```
alldata_agg <- with(alldata[alldata$both_applications_valid==1,],
                    aggregate(job_id,
                              list(coast=coast, good_resume=good_resume,
                                   call_back_factor=call_back_factor,
                                   university_brand=university_brand), length))
#rename stuff for aesthetics
alldata_agg$call_back_factor <- factor(alldata_agg$call_back_factor,
                                       levels=c('None', 'Rejection', 'Interview'))
alldata_agg$good_resume <- ifelse(alldata_agg$good_resume==1, "Strong Resume",
                                  "Weak Resume")
alldata_agg$university_brand<-ifelse(alldata_agg$university_brand==1, "Brand School",
                                  "Non-Brand School")

ggp <- ggplot(alldata_agg, aes(x=university_brand, y=x,
                               group=call_back_factor, fill=call_back_factor))
ggp + geom_bar(stat="identity", color="black", position=dodge)+
  facet_grid(good_resume~coast)+ylab('Number of responses')+
  ggtitle('Application responses by coast,\nbrand, and strength of resume')+
  theme_update(plot.title = element_text(hjust = 0.5))
```




```
alldata_agg <- with(alldata[alldata$both_applications_valid==1,],
                    aggregate(job_id,
                              list(coast=coast, size_bin=size_bin,
                                   call_back_factor=call_back_factor,
                                   university_brand=university_brand), length))
#rename stuff for aesthetics
alldata_agg$call_back_factor <- factor(alldata_agg$call_back_factor,
                                       levels=c('None', 'Rejection', 'Interview'))
alldata_agg$university_brand<-ifelse(alldata_agg$university_brand==1, "Brand School",
                                     "Non-Brand School")

ggp <- ggplot(alldata_agg, aes(x=university_brand, y=x,
                               group=call_back_factor, fill=call_back_factor))
ggp + geom_bar(stat="identity", color="black", position=dodge)+
  facet_grid(size_bin~coast)+ylab('Number of responses')+
  ggtitle('Application responses by coast,\nbrand, and company size')+
  theme_update(plot.title = element_text(hjust = 0.5))
```



Final model

```
lm.out <- lm(call_back_binary ~ coast+ phase+
               size_bin+staggered_application + university_brand*good_resume, data=alldata[alldata$both
summary(lm.out)
```

```
##
## Call:
## lm(formula = call_back_binary ~ coast + phase + size_bin + staggered_application +
##     university_brand * good_resume, data = alldata[alldata$both_applications_valid ==
##     1, ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30274 -0.13236 -0.06441  0.00096  0.92640
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   0.05779    0.06017   0.960  0.33797
## coastWest                    -0.10943    0.03946  -2.773  0.00608 **
## phasePhase2                  -0.03490    0.03985  -0.876  0.38218
## size_binMedium                0.09399    0.04551   2.065  0.04017 *
## size_binSmall                -0.02386    0.05132  -0.465  0.64250
## staggered_application         0.02315    0.03879   0.597  0.55139
## university_brand              0.07457    0.05503   1.355  0.17696
```

```
## good_resume                    0.12781    0.05508   2.320  0.02133 *
## university_brand:good_resume -0.16508    0.07850  -2.103  0.03671 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2784 on 201 degrees of freedom
## Multiple R-squared:  0.09818,    Adjusted R-squared:  0.06229
## F-statistic: 2.735 on 8 and 201 DF,  p-value: 0.006943
```

```
se.model1 = sqrt(diag(vcovHC(lm.out)))
stargazer(lm.out,
          se=list(se.model1),
          star.cutoffs=c(0.05, 0.01, 0.001), title = "Final Model",
          type="text")
```

```
##
## Final Model
## ===============================================================
##                                     Dependent variable:
##                                 ----------------------------
##                                       call_back_binary
## ---------------------------------------------------------------
## coastWest                                    -0.109**
##                                               (0.039)
##
## phasePhase2                                   -0.035
##                                               (0.040)
##
## size_binMedium                                0.094*
##                                               (0.046)
##
## size_binSmall                                 -0.024
##                                               (0.045)
##
## staggered_application                          0.023
##                                               (0.036)
##
## university_brand                               0.075
##                                               (0.050)
##
## good_resume                                   0.128*
##                                               (0.057)
##
## university_brand:good_resume                 -0.165*
##                                               (0.081)
##
## Constant                                       0.058
##                                               (0.050)
##
## ---------------------------------------------------------------
## Observations                                    210
## R2                                             0.098
## Adjusted R2                                    0.062
## Residual Std. Error                      0.278 (df = 201)
## F Statistic                            2.735** (df = 8; 201)
```

```
## ============================================================
## Note:                         *p<0.05; **p<0.01; ***p<0.001
```