Class Review

Professor Wehde

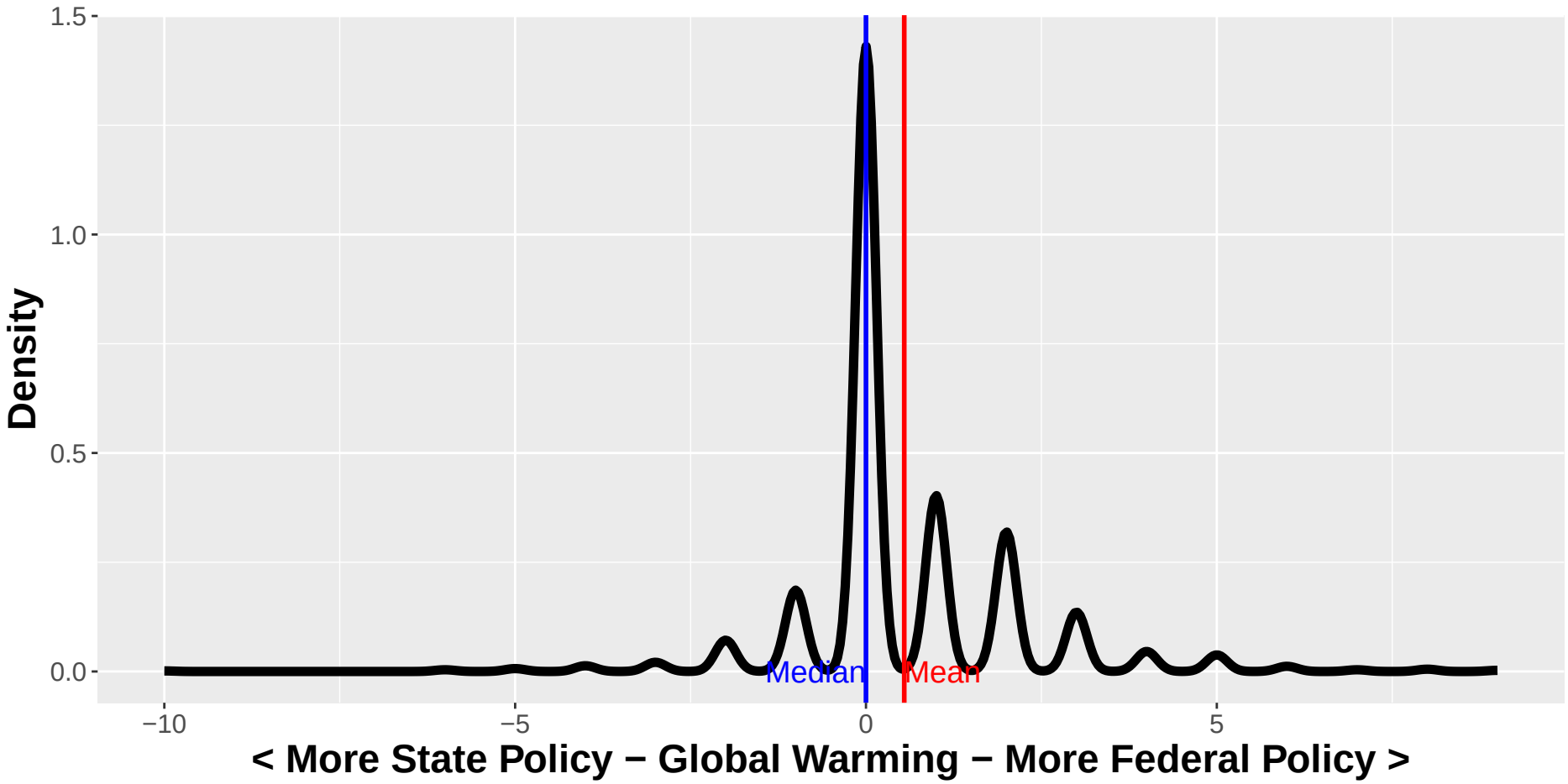
In class, we covered...

0. Descriptive Statistics and Visualizations
1. Two-sample t-test (difference of means)
   - DV: continuous variable
   - IV: dichotomous categorical variable
2. Two-proportion z-test (difference of proportions)
   - DV: dichotomous categorical variable
   - IV: dichotomous categorical variable
3. Covariance and correlation
   - DV: continuous variable
   - IV: continuous variable
4. Bivariate linear (simple) regression
   - DV: continuous variable
   - IV: continuous variable
5. Multiple linear regression
   - DV: continuous variable
   - IV: 2+ variables of any type
   - IV: categorical variable(s) and interaction terms

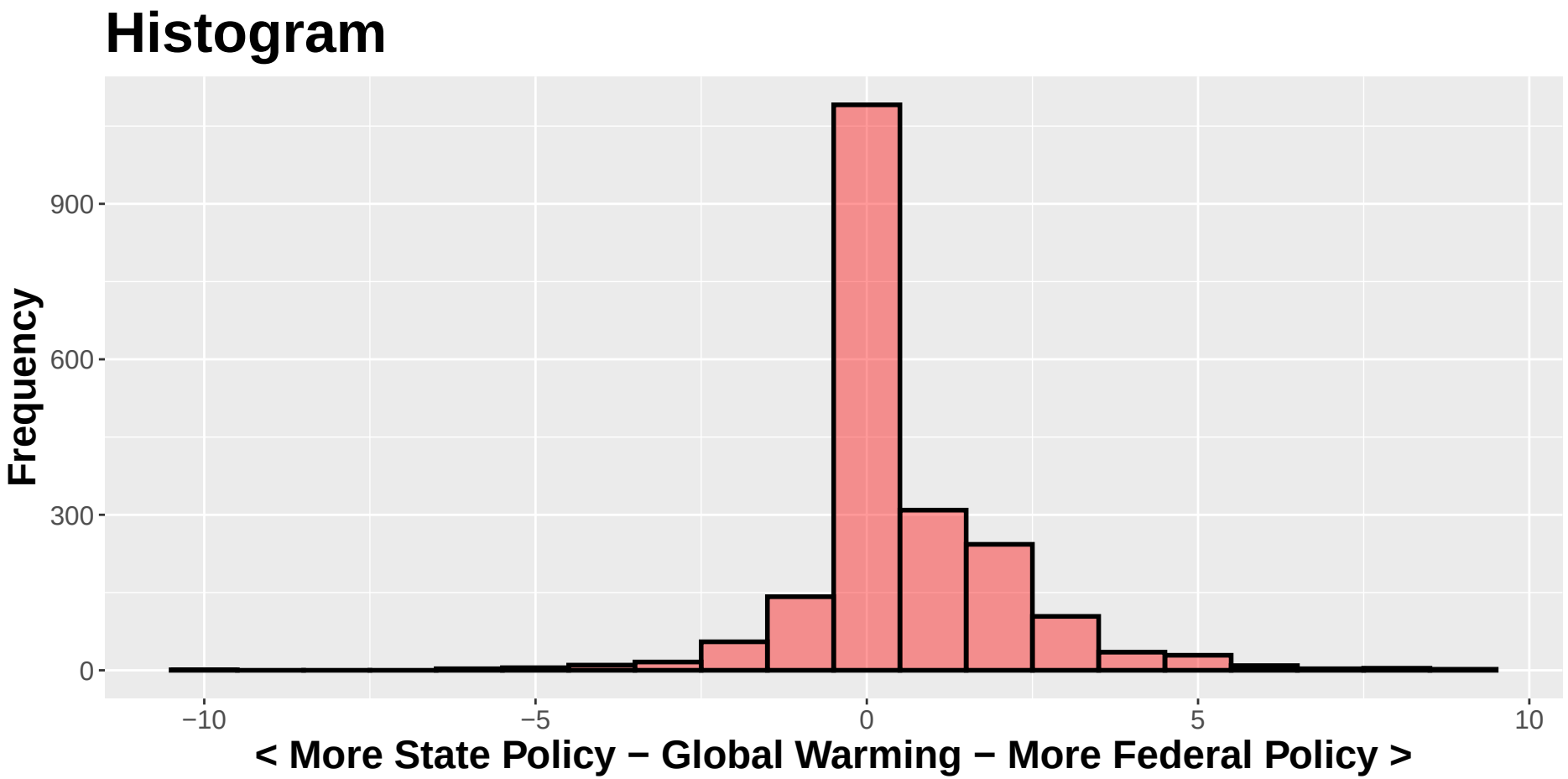
0. Descriptive Statistics and Visualizations

Depend on the type of data you have: - Can visualize entire distribution of continuous data using histograms or density pltos - Can visualize categorical differences in that distribution using a boxplot

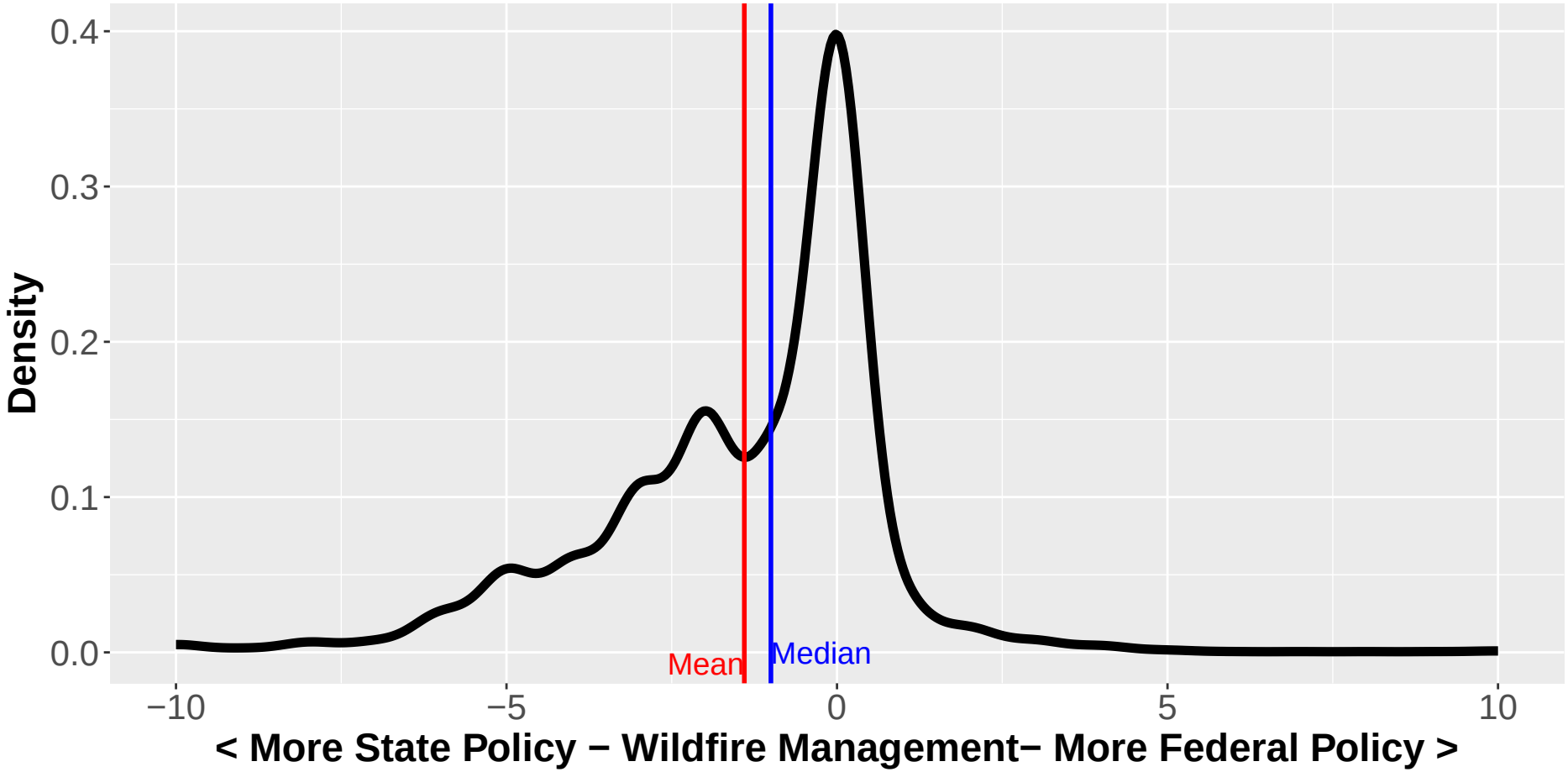
Density Plots



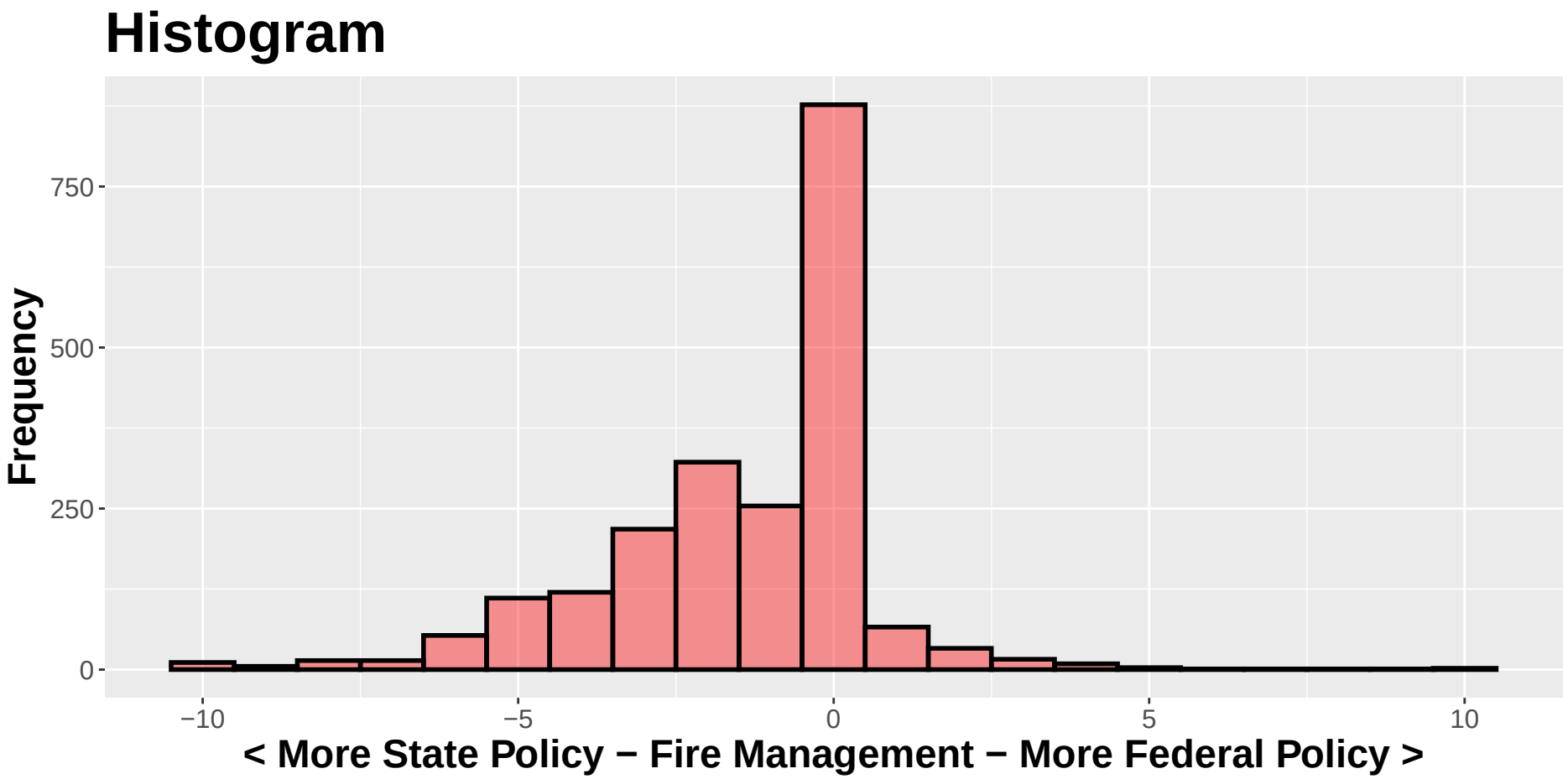
And Histograms



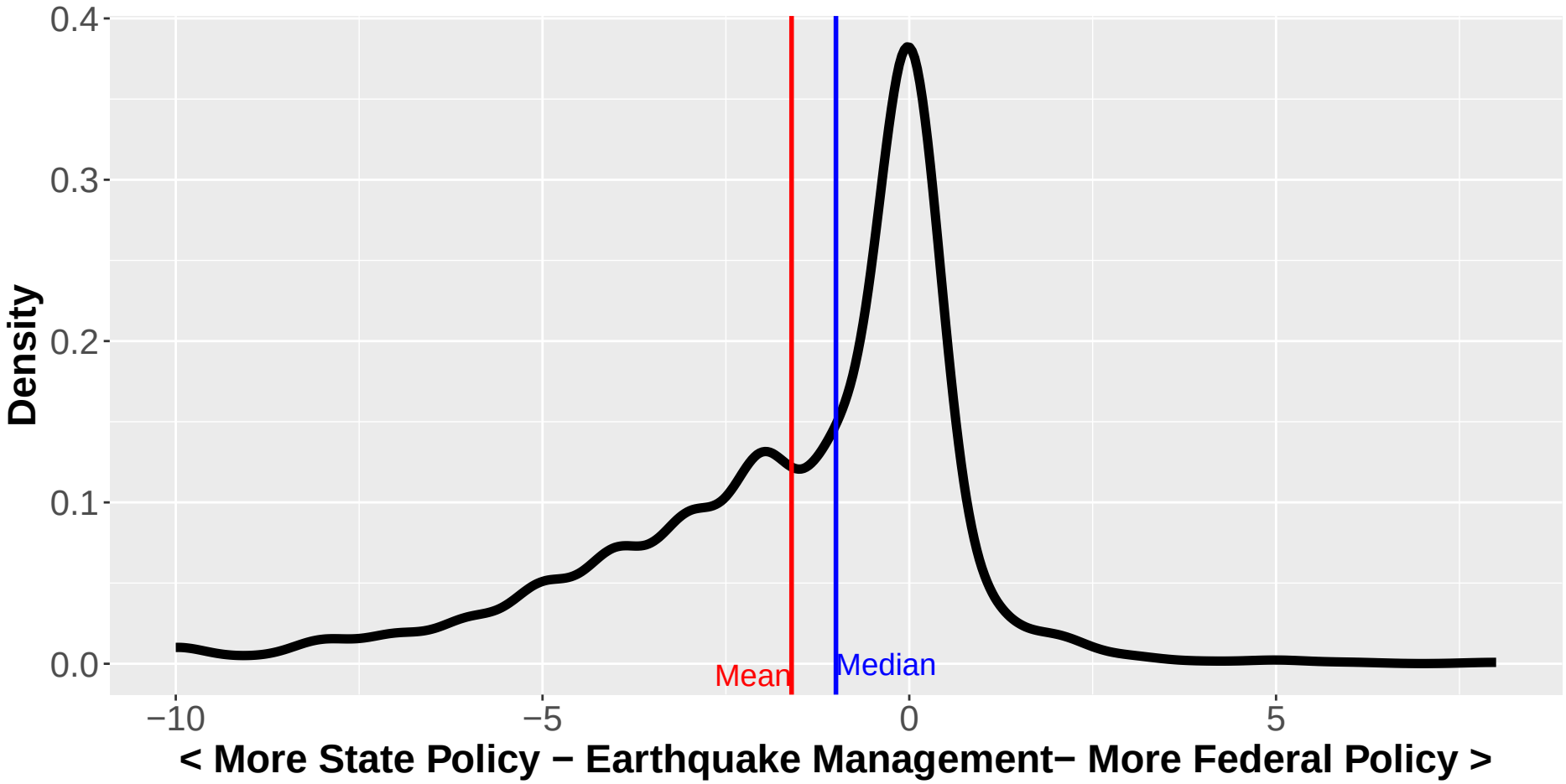
Density Plot



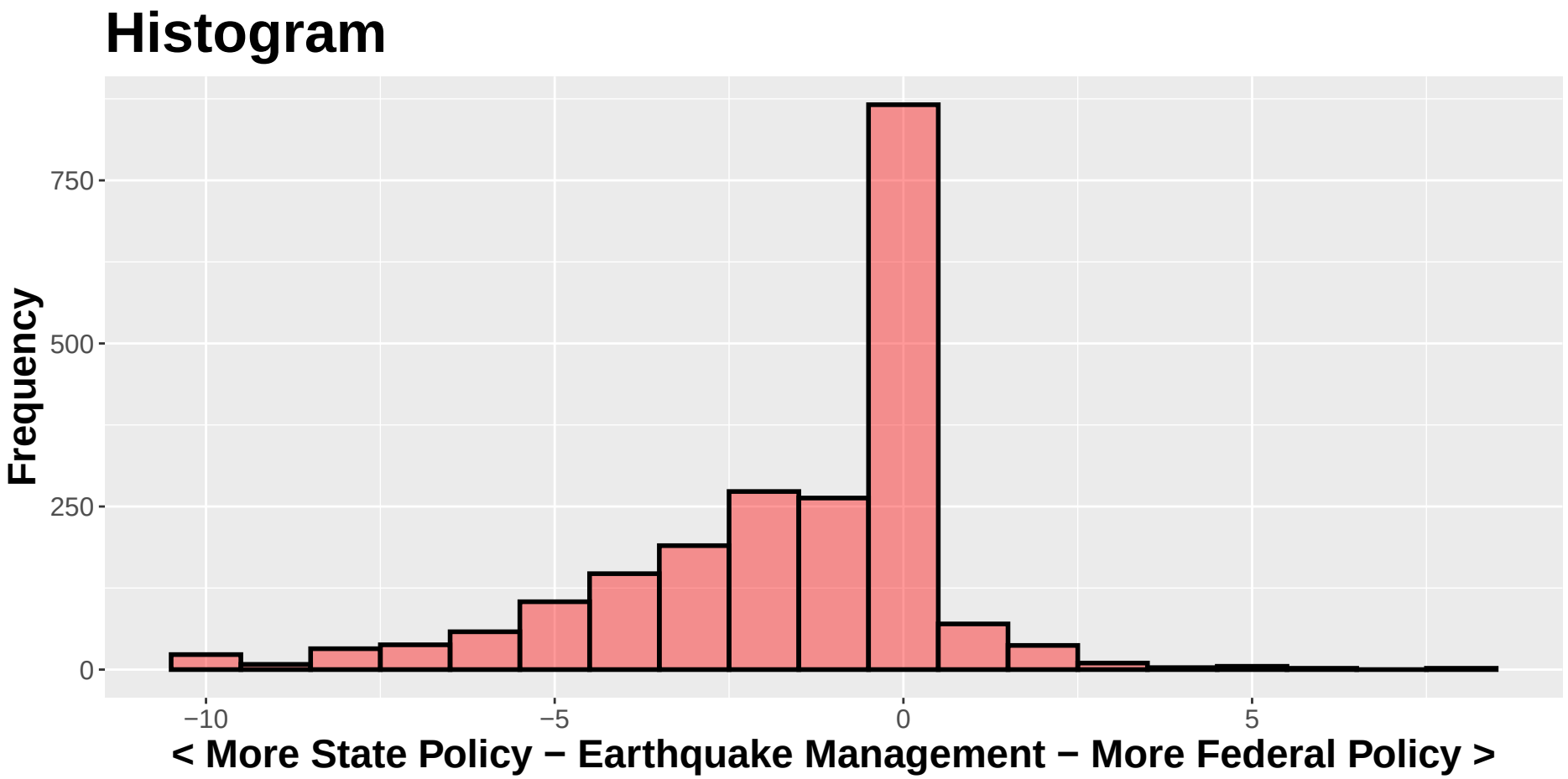
And Histograms



Density Plot



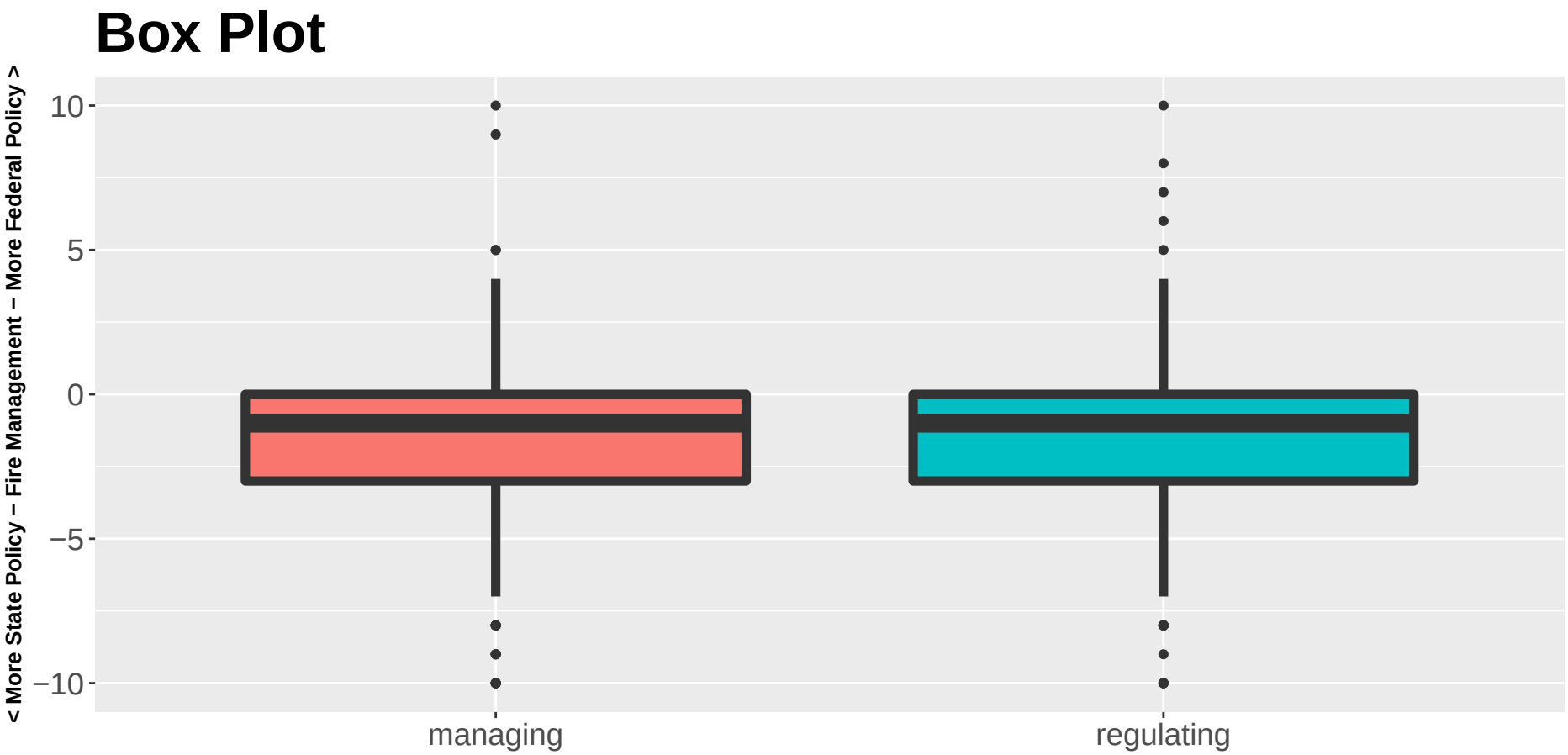
And Histograms



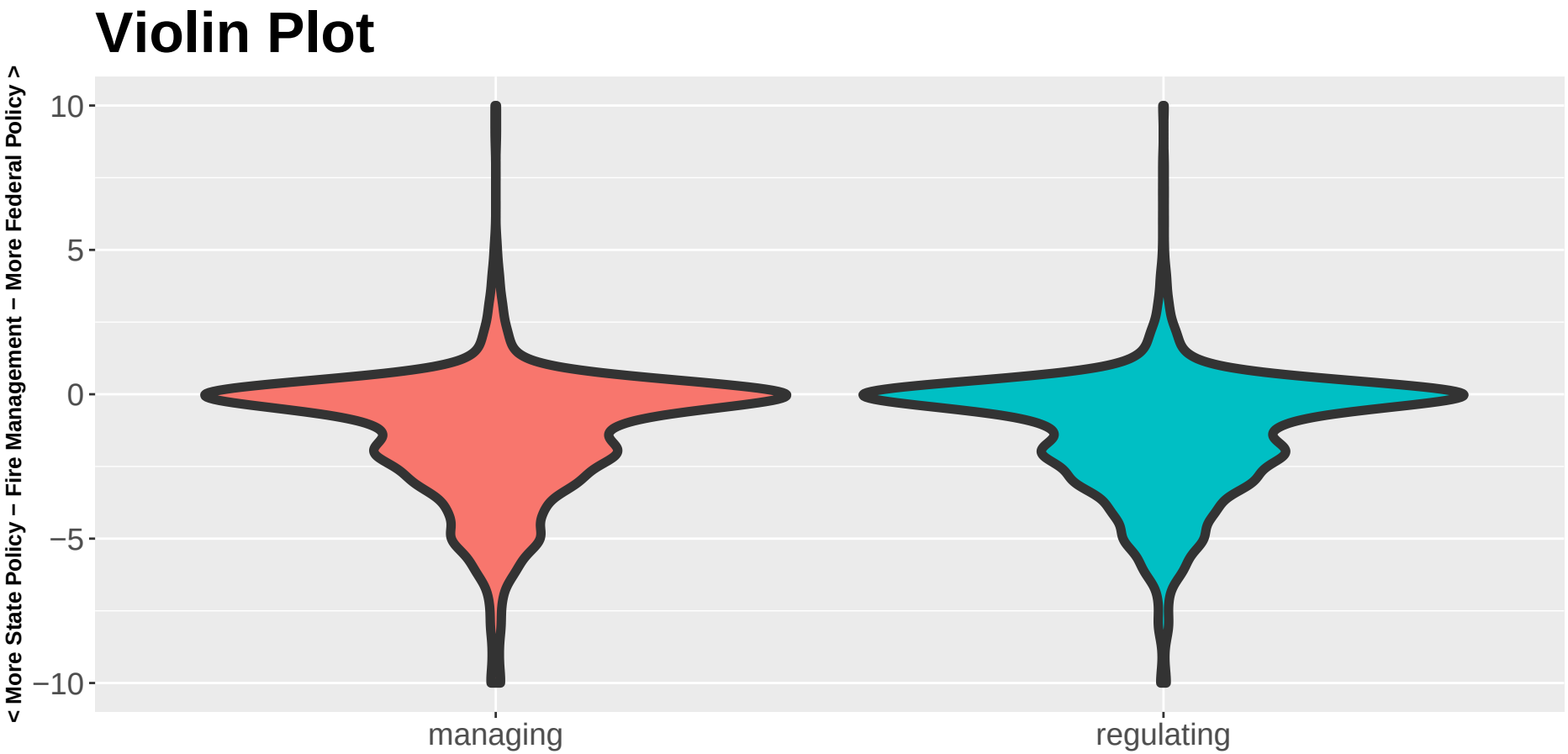
Visualizations for continuous DV and Categorical IV

- Boxplots
- Violin Plots

Boxplots

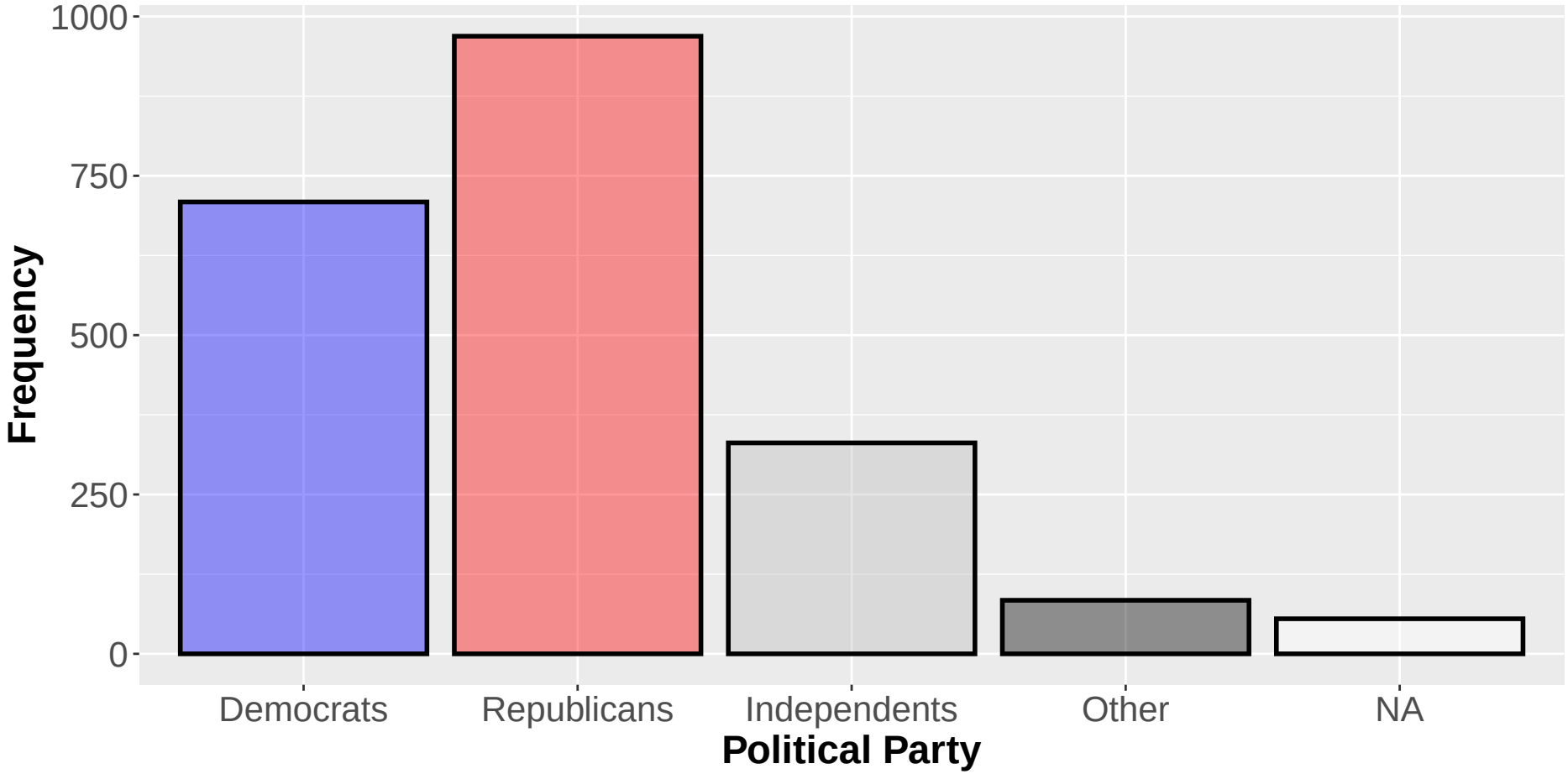

Box Plot
< More State Policy – Fire Management – More Federal Policy >
10
5
0
−5
−10
managing
regulating

Violin Plots



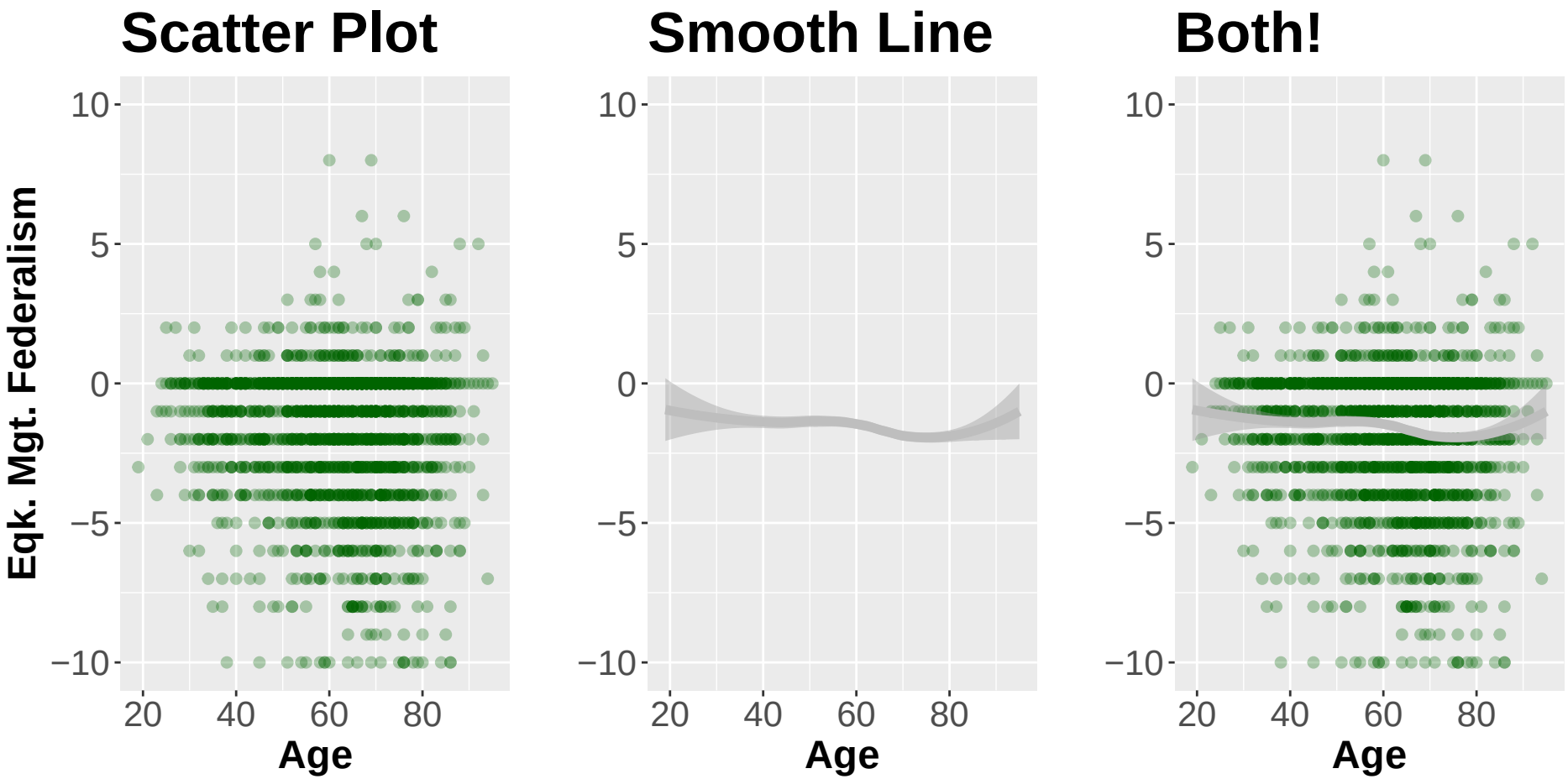
Visualizations for Categorical Variables.

- Bar plots

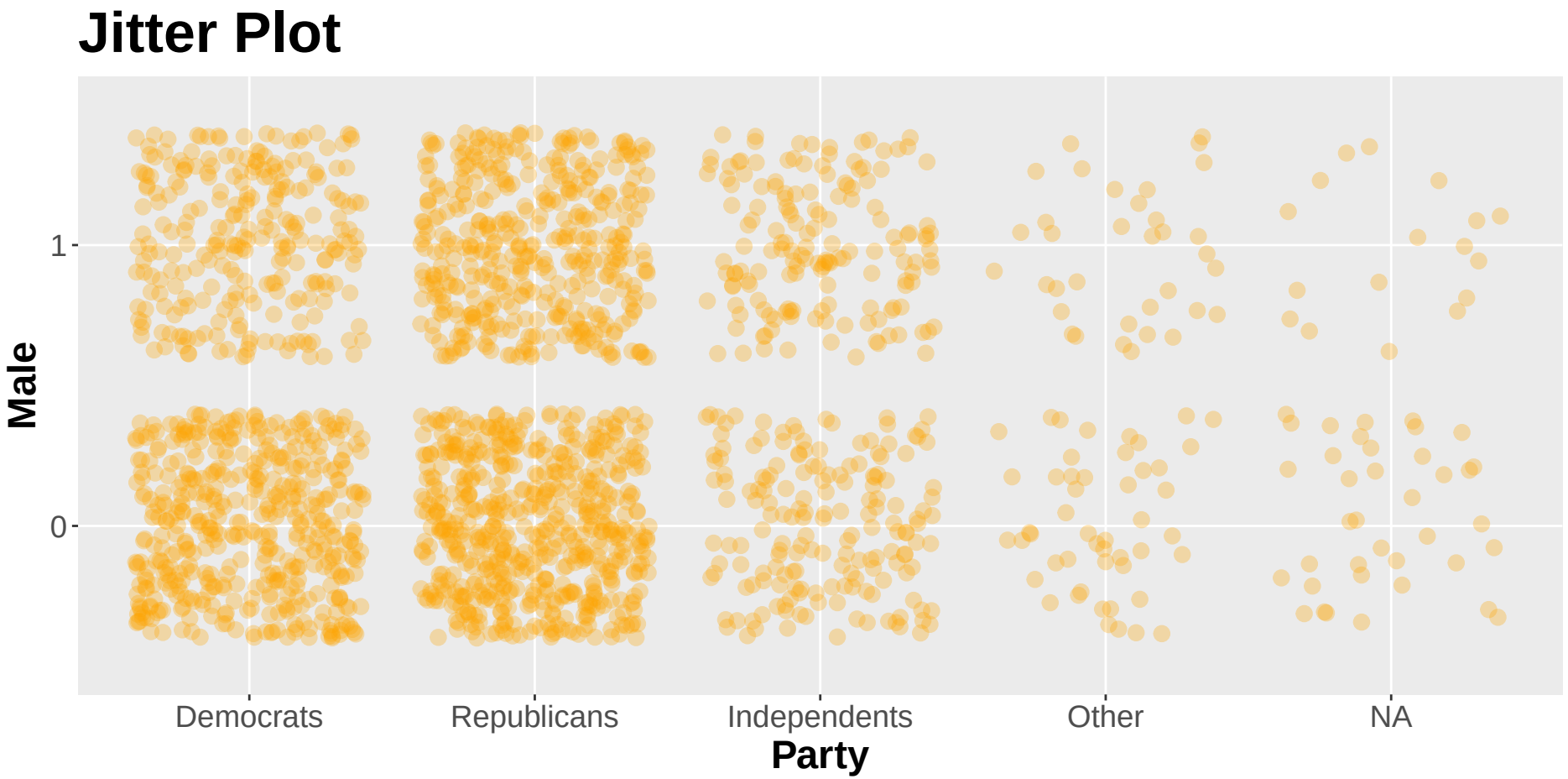



```
## $title
```

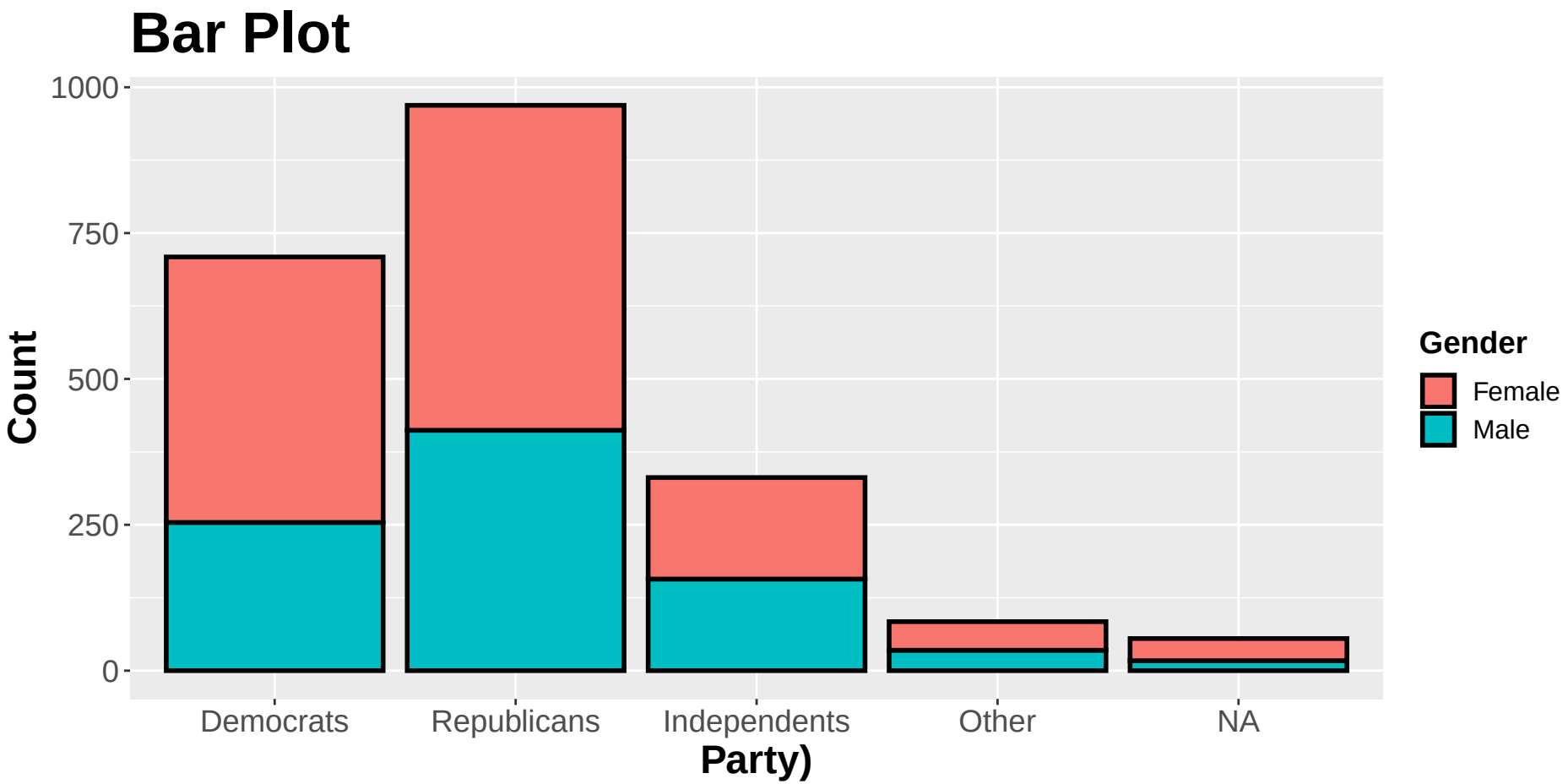
Visualizations of Continuous DV and IV



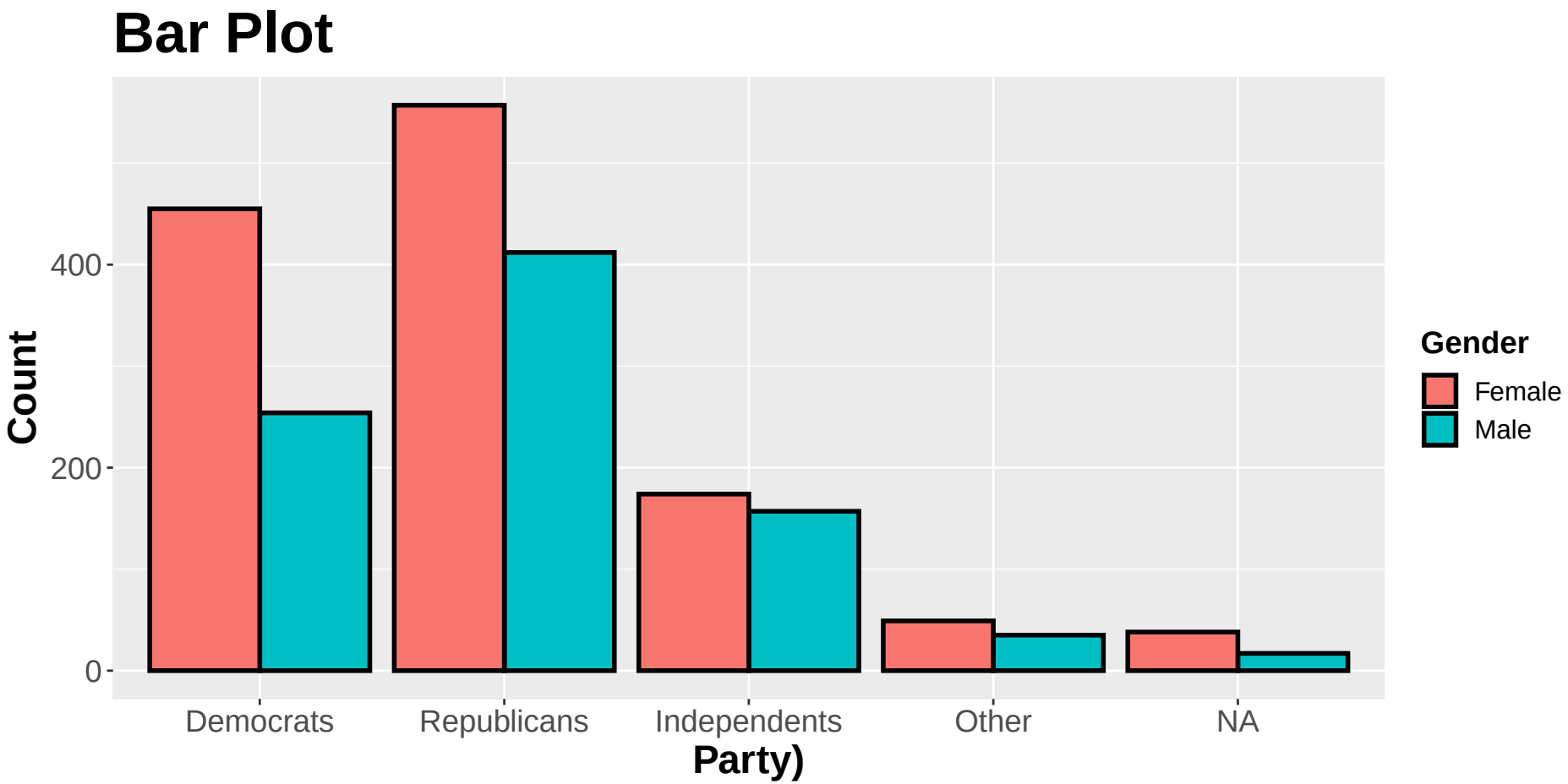
Visualizing Discrete IV and Discrete DV Variables



Visualizing Discrete IV and Discrete DV Variables



Visualizing Discrete IV and Discrete DV Variables



Descriptive Statistics in Tables

```
describe(wave1718$glbfed)
```

```
##    vars    n mean   sd median trimmed mad min max range skew kurtosis   se
## X1    1 2061 0.55 1.52      0    0.44   0 -10   9    19 0.73     5.09 0.03
```

Statistical Tests 0.5. 1 sample t-test Comparing Mean to value

```
t.test(wave1718$eqkfed, mu = 0)
```

```
##
##  One Sample t-test
##
## data:  wave1718$eqkfed
## t = -31.013, df = 2130, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -1.706364 -1.503397
## sample estimates:
## mean of x
##  -1.60488
```

Statistical Tests 1.Two-sample t-test (difference of means)

- DV: continuous variable
- IV: dichotomous categorical variable

```
t.test(wave1718$ethqk_risk_mgmt ~ wave1718$ethqk_risk_rand)
```

```
##
##  Welch Two Sample t-test
##
## data:  wave1718$ethqk_risk_mgmt by wave1718$ethqk_risk_rand
## t = -0.8397, df = 2131, p-value = 0.4012
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3245430  0.1299415
## sample estimates:
##   mean in group managing mean in group regulating
##                 7.023629                 7.120930
```

Statistical Tests 1.Two-sample t-test (difference of means)

```
t.test(wave1718$fire_risk_fed_mgmt ~ wave1718$fire_risk_rand)
```

```
##
##  Welch Two Sample t-test
##
## data:  wave1718$fire_risk_fed_mgmt by wave1718$fire_risk_rand
## t = 2.609, df = 2139.6, p-value = 0.009143
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.08235799 0.58088618
## sample estimates:
##   mean in group managing mean in group regulating
##                 5.612754                 5.281132
```

Statistical Tests 2. Two-proportion z-test (difference of proportions)

- DV: dichotomous categorical variable
- IV: dichotomous categorical variable

```
prop.test(table(wave1718$glbcc, wave1718$gender), correct=FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  table(wave1718$glbcc, wave1718$gender)
## X-squared = 12.629, df = 1, p-value = 0.0003798
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.11792484 -0.03405262
## sample estimates:
##    prop 1    prop 2
## 0.5497326 0.6257214
```

Statistical Tests 3. Covariance and correlation

- DV: continuous variable
- IV: continuous variable

```
cov(wave1718$age, wave1718$glbfed, use = "complete.obs")
```

```
## [1] -1.286061
```

```
cor(wave1718$age, wave1718$glbfed, use = "complete.obs")
```

```
## [1] -0.06139664
```

Statistical Tests 3. Covariance and correlation

```
corr.test(wave1718$age, wave1718$glbfed, use = "complete.obs")
```

```
## Call:corr.test(x = wave1718$age, y = wave1718$glbfed, use = "complete.obs")
## Correlation matrix
## [1] -0.06
## Sample Size
## [1] 2061
## Probability values  adjusted for multiple tests.
## [1] 0.01
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

Statistical Tests 4. Bivariate linear (simple) regression

- DV: continuous variable
- IV: continuous variable

```
mod1<-lm(glbfed ~ age, data = wave1718)

summary(mod1)
```

```
##
## Call:
## lm(formula = glbfed ~ age, data = wave1718)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.3988  -0.5891  -0.4803   0.5129   8.5061
##
## Coefficients:
##              Estimate Std. Error t value       Pr(>|t|)
## (Intercept)  0.962911   0.153295   6.281 0.000000000408 ***
## age         -0.006797   0.002435  -2.791         0.0053 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.52 on 2059 degrees of freedom
##   (87 observations deleted due to missingness)
## Multiple R-squared:  0.00377,    Adjusted R-squared:  0.003286
## F-statistic: 7.791 on 1 and 2059 DF,  p-value: 0.0053
```

Statistical Tests 5. Multiple linear regression

- DV: continuous variable
- IV: 2+ variables of any type
- IV: categorical variable(s) and interaction terms

```
vensglb<-lm(glbfed ~ age + f.gender + bach + grad + income + rep + ideol.x
            + white + as.factor(glbwrm_ok) +
              as.factor(glbcc) + glbcc_risk , wave1718)

summary(vensglb)
```

Statistical Tests 5. Multiple linear regression

```
##
## Call:
## lm(formula = glbfed ~ age + f.gender + bach + grad + income +
##     rep + ideol.x + white + as.factor(glbwrm_ok) + as.factor(glbcc) +
##     glbcc_risk, data = wave1718)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.0209  -0.7990  -0.2301   0.5801   8.1084
##
## Coefficients:
##                          Estimate Std. Error t value    Pr(>|t|)
## (Intercept)             -0.679733   0.650122  -1.046    0.295908
## age                     -0.004705   0.002621  -1.795    0.072804 .
## f.gender1                0.233593   0.071659   3.260    0.001135 **
## bach1                    0.322864   0.085017   3.798    0.000151 ***
## grad1                    0.088909   0.093539   0.950    0.341987
## income                   0.088162   0.054053   1.631    0.103057
## rep1                     0.026296   0.087998   0.299    0.765109
## ideol.x                 -0.049744   0.029127  -1.708    0.087834 .
## white1                   0.165435   0.110764   1.494    0.135455
## as.factor(glbwrm_ok)1    0.056310   0.140202   0.402    0.687999
## as.factor(glbwrm_ok)2    0.378093   0.109290   3.460    0.000553 ***
## as.factor(glbcc)1        0.575626   0.110745   5.198 0.000000224 ***
## glbcc_risk              -0.006949   0.019788  -0.351    0.725503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.489 on 1840 degrees of freedom
##   (295 observations deleted due to missingness)
## Multiple R-squared:  0.07084,    Adjusted R-squared:  0.06478
## F-statistic: 11.69 on 12 and 1840 DF,  p-value: < 0.00000000000000022
```

Statistical Tests 5. Multiple linear regression