POLS 7012 Final Exam (Answer Key)

YOUR NAME HERE
University of Georgia
email@email.email

Joseph T. Ornstein
University of Georgia
jornstein@uga.edu

December 09, 2020

1 Introduction

In this paper, we will replicate the results from "Civic Honesty Around the Globe" (Cohn et al. 2019). In that paper, the research team dropped literally thousands of wallets in cities across the globe to see how many of them get returned. They find, interestingly, that wallets with money in them are *more* likely to be returned than wallets without money, which is all very nice and pleasant.

Before you begin, skim over the paper and its Supplementary Materials, included in the `papers/` folder. To replicate the findings, follow the instructions in each section below, writing your code in the chunks provided. Anything I call an "extra challenge" is available to you if you'd like to give it a try, but is not required. To submit your final exam, knit this `.Rmd` to a PDF and post both files (`.Rmd` and PDF) to eLC.

2 Data

Replication files are available here, and I have taken the liberty of downloading them into the `data/` folder. The dataset we'll need for this replication is called the behavioral data. Let's start by loading it.

```
data <- read_csv('data/behavioral data (csv file).csv')
```

3 Results

3.1 Replicating Figure 1

Our first task is to replicate the left-hand side of Figure 1. To do so, we need to perform the following steps:

- Keep only the observations in the Money and NoMoney conditions.
- Recode the `cond` variable as "Money" and "NoMoney".
- Compute the average reporting rate, grouped by country and condition.
- Build a scatter plot with average reporting rate on the x-axis, country on the y-axis, and colored by monetary condition.

As an extra challenge, you can do any combination of the following:

- Rearrange the y-axis so that the countries with the lowest reporting rate appear at the bottom and those with the highest reporting rate appear at the top.
- Include line segments between points as in the original figure
- Use the colors from the original figure
- Use `geom_text()` to add floating labels for each country like in the original figure. This is an extra, *extra* challenge, but man is it satisfying when you get it right!

```
# Here's the basic version
fig1 <- data %>%
  filter(cond %in% c(0,1)) %>%
  mutate(cond = case_when(cond == 1 ~ 'Money',
                          cond == 0 ~ 'NoMoney')) %>%
  # compute reporting rate by country and monetary condition
  group_by(Country,
           cond) %>%
  summarize(pct_reported = mean(response)) %>%
  # plot
  ggplot() +
  geom_point(aes(x=pct_reported, y=Country, color = factor(cond))) +
  labs(x = 'Reporting rate (%)', y = '', color = 'Condition') +
  theme_minimal()

fig1
```

```r
# Here's the challenging version
fig1a <- data %>%
  filter(cond %in% c(0,1)) %>%
  mutate(cond = case_when(cond == 1 ~ 'Money',
                          cond == 0 ~ 'NoMoney')) %>%
  # compute reporting rate by country and monetary condition
  group_by(Country,
           cond) %>%
  summarize(pct_reported = mean(response)) %>%
  # pivot_wider to make those line segments
  ungroup %>%
  pivot_wider(names_from = cond, values_from = pct_reported) %>%
  # reorder Country by the NoMoney reporting rate
  mutate(Country = fct_reorder(Country, NoMoney)) %>%
  # compute label position, left of the minimum reporting rate
  mutate(label_position = pmin(Money, NoMoney) -
           nchar(as.character(Country))/3.5 - 1) %>%
  # begin ggplot
  ggplot() +
  geom_segment(aes(x=Money, xend=NoMoney, y=Country, yend=Country),
               color = 'gray', size = 0.5) +
  geom_point(aes(x=Money,y=Country), color = 'red') +
  geom_point(aes(x=NoMoney,y=Country), color = '#F6BE00') +
  geom_text(aes(x=label_position, y=Country, label = Country), size = 2) +
  labs(x = 'Reporting rate (%)', y = '', color = 'Condition') +
  theme_classic() +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.line.y = element_blank())

fig1a
```

3.2 Replicating Figure 2

Now replicate Figure 2. To do so, we need to perform the following steps:

- Keep only the data from Poland, the United Kingdom, and the United States
- Keep only the NoMoney, Money, and BigMoney conditions
- Recode the `cond` variable as "NoMoney", "Money" and "BigMoney"

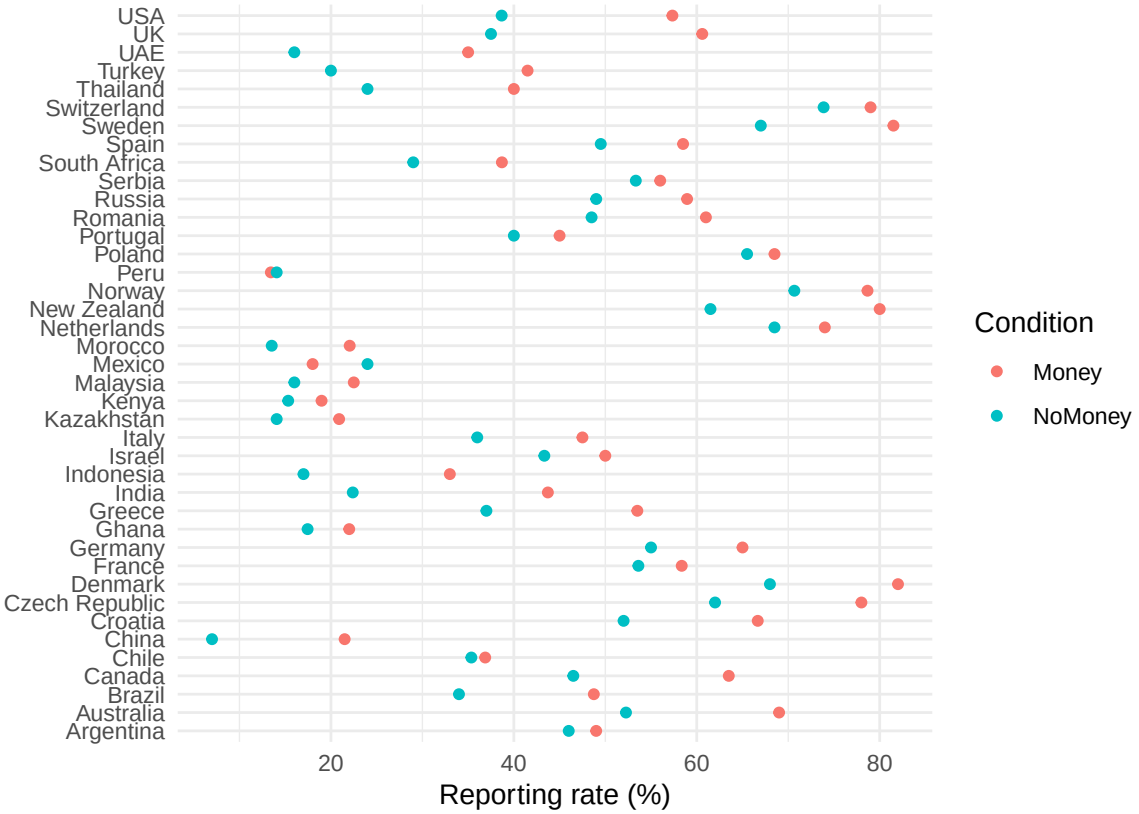

Figure 1: Share of wallets reported in the NoMoney and Money conditions, by country.

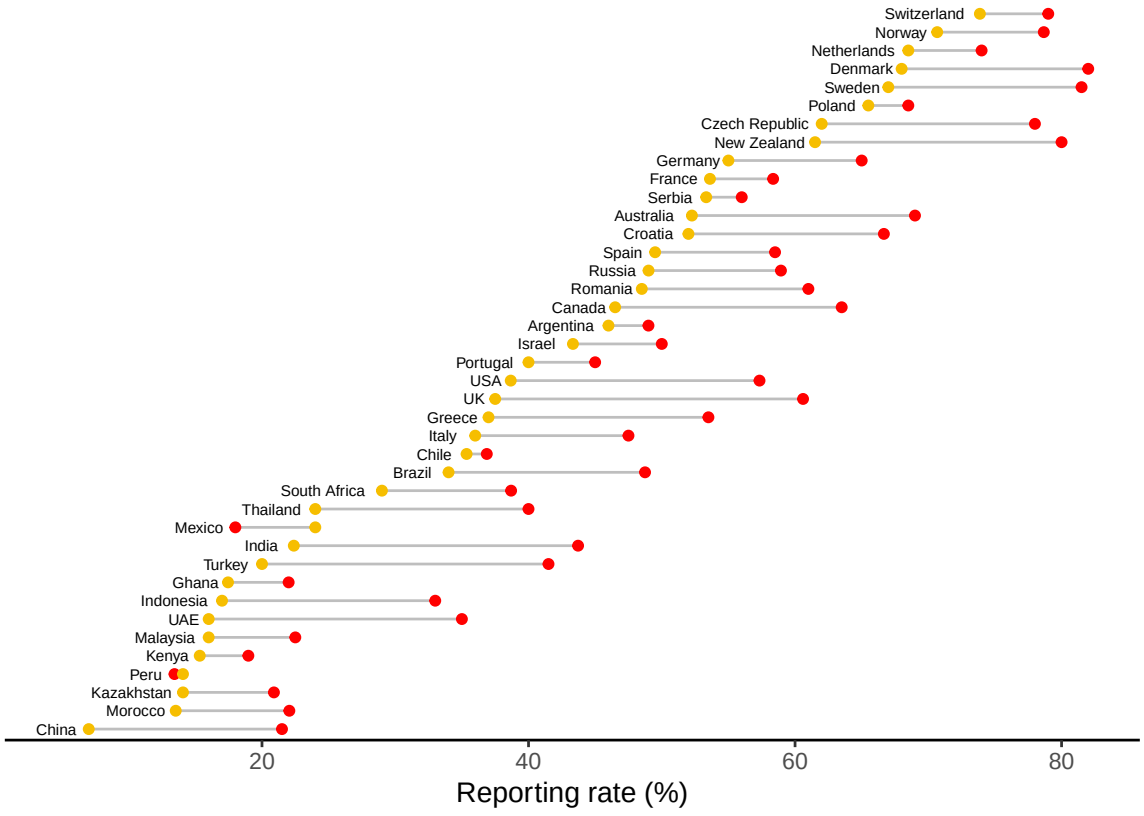

Figure 2: Share of wallets reported in the NoMoney and Money conditions, by country.

- Compute the average response rate, grouped by country and condition.
- Plot a scatter with condition on the x-axis, reporting rate on the y-axis, and colored by country.
- Add a `geom_line()` layer with the same aesthetics (also include `group = Country` as an aesthetic).

As an extra challenge, you can do any combination of the following:

- Use original colors from the paper
- Use the ggplot theme that best matches the theme from the paper
- Reorder the `cond` variable so it appears in the same order as the original Figure 2
- Include country labels as in the paper with `geom_dl()` from the `directlabels` package, and remove the legend

```
# Basic version
data %>%
  filter(Country %in% c('Poland', 'UK', 'USA'),
         cond %in% 0:2) %>%
  mutate(cond = case_when(cond == 0 ~ 'NoMoney',
                          cond == 1 ~ 'Money',
                          cond == 2 ~ 'BigMoney')) %>%
  group_by(Country, cond) %>%
  summarize(reporting_rate = mean(response)) %>%
  ggplot() +
  geom_point(mapping = aes(x = cond, y = reporting_rate,
                           color = Country)) +
  geom_line(mapping = aes(x = cond, y = reporting_rate,
                          group = Country, color = Country)) +
  labs(x = '', y = 'Reporting Rate (%)')
```

```
library(directlabels)

# Harder version
data %>%
  filter(Country %in% c('Poland', 'UK', 'USA'),
         cond %in% 0:2) %>%
  mutate(Country = case_when(Country == 'UK' ~ 'United Kingdom',
                             Country == 'USA' ~ 'United States',
                             TRUE ~ Country),
```

```
        cond = case_when(cond == 0 ~ 'NoMoney',
                         cond == 1 ~ 'Money',
                         cond == 2 ~ 'BigMoney')) %>%
  group_by(Country, cond) %>%
  summarize(reporting_rate = mean(response)) %>%
  mutate(cond = factor(cond, levels = c('NoMoney', 'Money', 'BigMoney'))) %>%
  ggplot() +
  geom_point(mapping = aes(x = cond, y = reporting_rate,
                           color = Country)) +
  geom_line(mapping = aes(x = cond, y = reporting_rate,
                          group = Country, color = Country)) +
  geom_dl(mapping = aes(x=cond, y=reporting_rate, label = Country),
          method = list(dl.trans(x = x + 0.2), 'last.points', cex = 0.75)) +
  scale_color_manual(values = c('#8b1b1e', '#f23524', '#f98f1c')) +
  labs(x = '', y = 'Reporting Rate (%)') +
  theme_classic() +
  theme(legend.position = 'none')
```

3.3 The Effect of Money

In paragraph 3 of page 2, the authors claim that citizens are "overwhelmingly more likely to report lost wallets containing money than those without". Let's replicate that finding. To do so, we must:

- Keep the observations in the Money and NoMoney conditions
- Compute the percent of subjects that reported wallets without money. Save that value, rounded to the nearest percent, as the object `no_money_reporting_rate`.
- Compute the percent of subjects that reported wallets with money. Save that value, rounded to the nearest percent, as the object `money_reporting_rate`.

On average, adding money to the wallet increased the likelihood of being reported from 40% in the NoMoney condition to 51% in the Money condition.

Next, conduct a hypothesis test to determine the strength of that result. Our null hypothesis is that there is no difference in reporting rates between the Money and NoMoney groups. Note that the explanatory variable (money or no money) and outcome variable (report or not report) are both categorical. Report the p-value associated with your null hypothesis test.

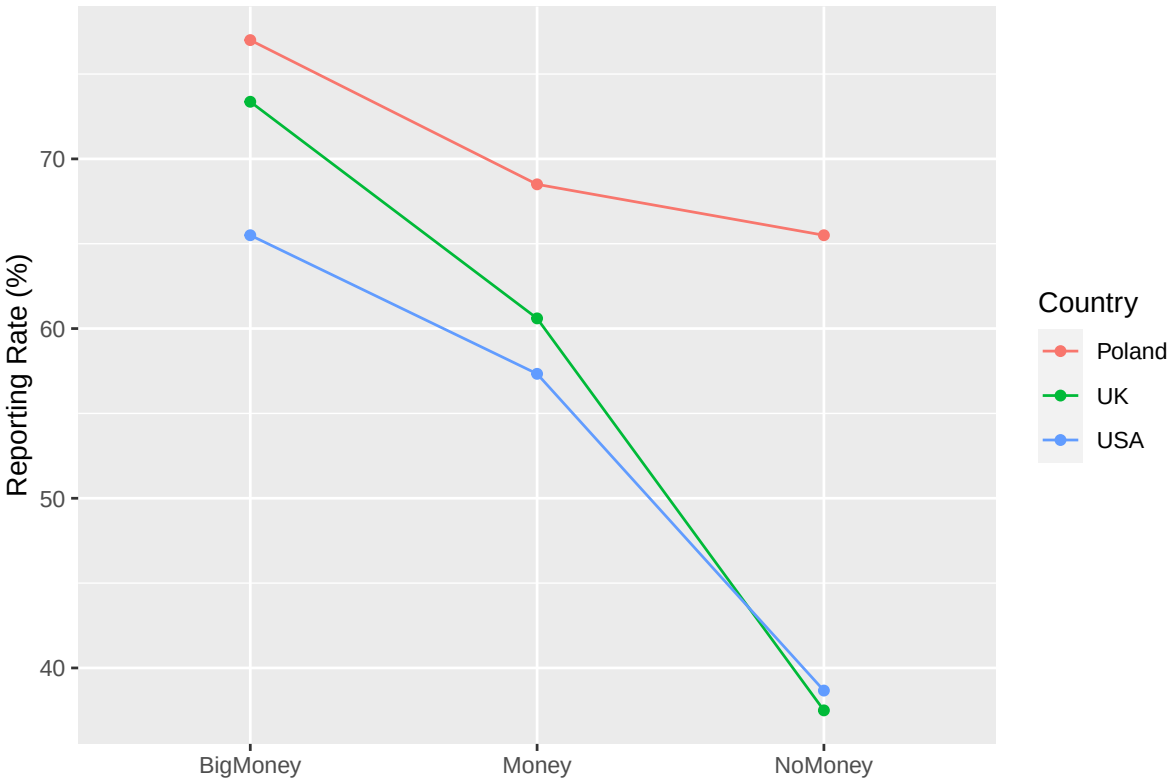

Figure 3: Reporting rates as a function of monetary stakes

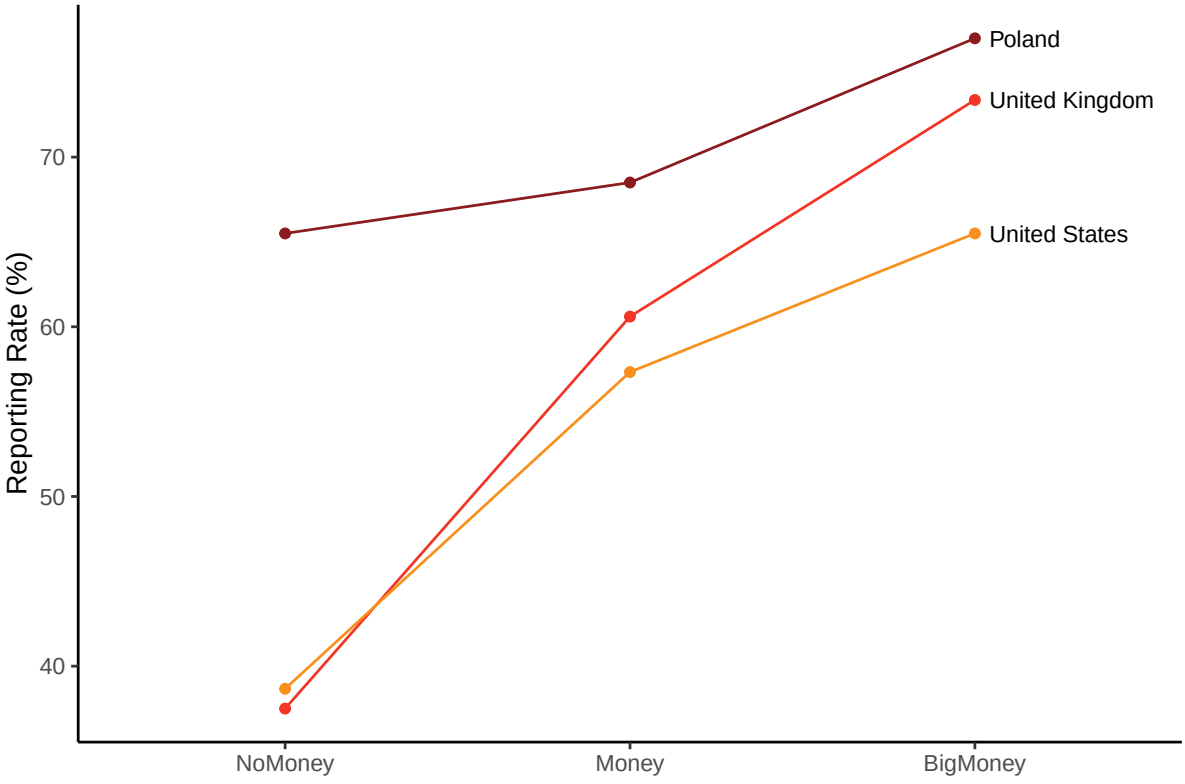

Figure 4: Reporting rates as a function of monetary stakes

```
data %>%
  filter(cond %in% c(0,1)) %>%
  select(cond, response) %>%
  table %>%
  chisq.test(correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  .
## X-squared = 201.21, df = 1, p-value < 2.2e-16
```

3.3.1 Replicating Table S8

In the Supplementary Materials, the authors extend the bivariate hypothesis test we just did by estimating two linear models that include other subject characterstics as covariates. To replicate that finding, we must:

- Create a variable called `Money`, equal to 1 if the subject's wallet contained money and 0 otherwise.
- Create a variable called `BigMoney`, equal to 1 if the subject's wallet contained **lots** of money and 0 otherwise.
- Create a variable called `MoneyNoKey`, equal to 1 if they subject's wallet contained money but no key and 0 otherwise.
- To include "Institution and City Fixed Effects", make sure that the city and type of institution are coded as characters or factors.
- Estimate your first model, `lm1`, including the treatment variables you created plus the institution and city fixed effects.
- In the second model `lm2`, add the binary variables mentioned in the notes of Table S8.
- Don't worry if your standard errors don't exactly match up. You'll learn more about "robust" standard errors next semester.

```
lm1 <- lm(response ~ male, data = data)
lm2 <- lm(response ~ male, data = data)

data <- data %>%
  mutate(Money = if_else(cond == 1, 1, 0),
         BigMoney = if_else(cond == 2, 1, 0),
         MoneyNoKey = if_else(cond == 3, 1, 0),
```

```
         institution = factor(institution),
         city = factor(city))

lm1 <- lm(response ~ Money + BigMoney + MoneyNoKey + institution + city, data = data)
lm2 <- lm(response ~ Money + BigMoney + MoneyNoKey + institution + city +
            male + above40 + computer + coworkers + other_bystanders, data = data)
```

The following code chunk outputs a pretty table like the one in Table S8. I provide it to you free of charge. All you need to do is make sure that `lm1` and `lm2` are correctly specified.

Table 1: Reporting rates in the Money and NoMoney condition

	(1)	(2)
Money	10.828***	10.791***
	(0.716)	(0.713)
male		−2.076**
		(0.740)
above40		−2.030**
		(0.737)
computer		6.874***
		(0.975)
coworkers		4.675***
		(0.771)
other_bystanders		−3.900***
		(0.798)
Constant	34.620**	33.302**
	(10.687)	(10.720)
Observations	17,303	17,295
Adjusted R^2	0.178	0.185
Note:	*p<0.05; **p<0.01; ***p<0.001	

4 Conclusion

Hey people are more likely to try to return wallets they find if they have money in them. Isn't that nice?

References

Cohn, Alain, Michel André Maréchal, David Tannenbaum, and Christian Lukas Zünd. 2019. "Civic Honesty Around the Globe." *Science* 365 (6448): 70–73. https://doi.org/10.1126/science.aau8712.