

CDS-101 Checkpoint #2

Eunho Cha, Daeun Choi, Songlee Jun, Hyun Woo Kang
Dawon Kyoung, Byungwook Oh, Duy Tran

2024-05-19

#GHG = Greenhouse Gases #FJO = Female Job Occupation

```
colnames(Group_dataset) <- c("Year", "Birth_GR", "GHG_GR", "Inflation_GR", "FJO_GR")  
  
view(Group_dataset)
```

#Summarize

```
summary(Group_dataset)
```

```
##      Year      Birth_GR      GHG_GR      Inflation_GR  
## Min.   :1991  Min.   :-5.870  Min.   :-17.080  Min.   :-6.7000  
## 1st Qu.:1998  1st Qu.: -3.555  1st Qu.:  1.065  1st Qu.: -0.9900  
## Median :2006  Median : -2.450  Median :  2.540  Median :  0.0800  
## Mean   :2006  Mean   : -2.499  Mean   :  3.181  Mean   : -0.1968  
## 3rd Qu.:2014  3rd Qu.: -1.305  3rd Qu.:  6.940  3rd Qu.:  0.8650  
## Max.   :2021  Max.   :  1.080  Max.   : 11.860  Max.   :  3.0700  
##      FJO_GR  
## Min.   : -7.000  
## 1st Qu.:  1.000  
## Median :  2.000  
## Mean   :  1.581  
## 3rd Qu.:  3.000  
## Max.   :  5.000
```

#Select

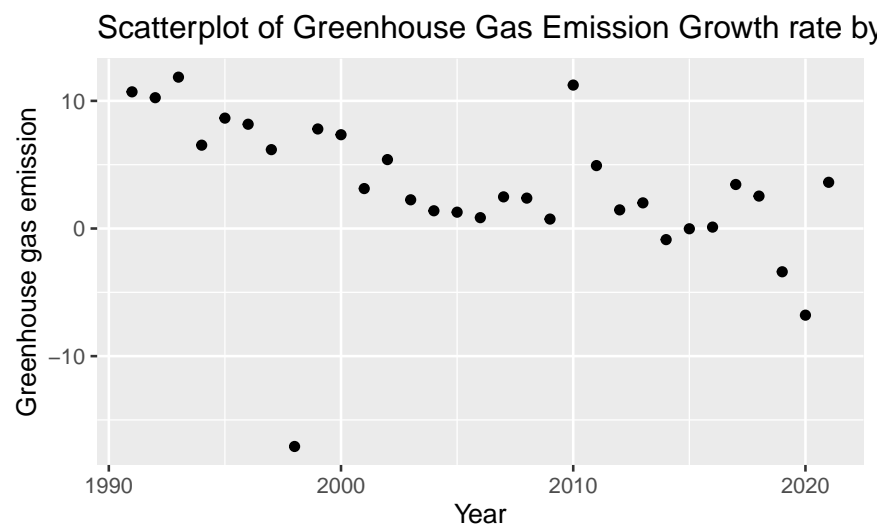
```
Envo_ft <- Group_dataset %>%  
  select(Year, Birth_GR, GHG_GR)
```

```
Econ_ft <- Group_dataset %>%  
  select(Year, Birth_GR, Inflation_GR)
```

```
Soci_ft <- Group_dataset %>%
  select(Year, Birth_GR, FJO_GR)
```

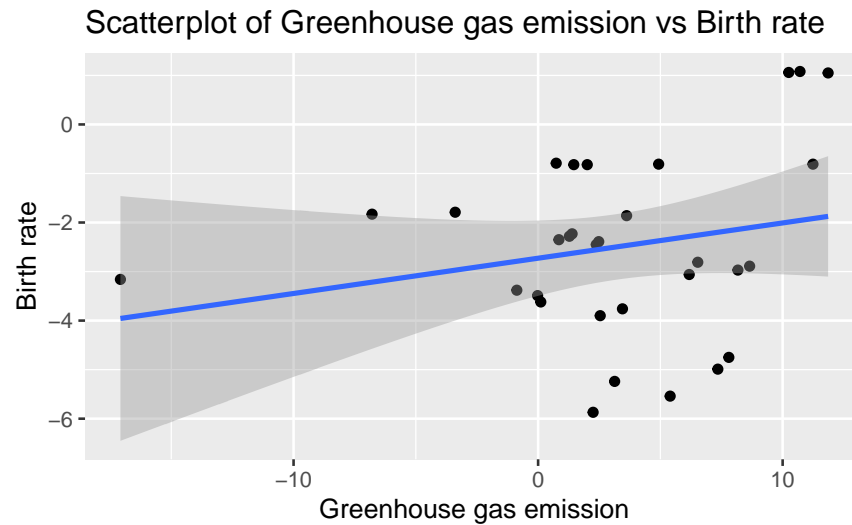
#Variation and Covariation - Envo_ft

```
Envo_ft %>%
  ggplot() +
  geom_point(mapping = aes(y = GHG_GR, x = Year)) +
  labs(title = "Scatterplot of Greenhouse Gas Emission Growth rate by Year",
       x = "Year",
       y = "Greenhouse gas emission")
```



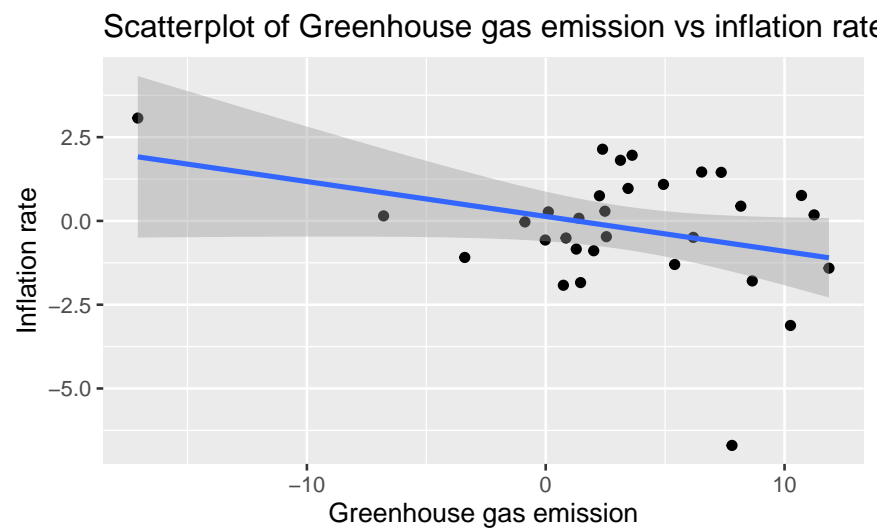
```
Envo_ft %>%
  ggplot()+
  geom_point(mapping = aes(x = GHG_GR, y = Birth_GR)) +
  geom_smooth(mapping = aes(x = GHG_GR, y = Birth_GR), method="lm")+
  labs(
    title = "Scatterplot of Greenhouse gas emission vs Birth rate",
    x = 'Greenhouse gas emission',
    y = "Birth rate")
```

'geom_smooth()' using formula = 'y ~ x'



```
Group_dataset %>%
  ggplot() +
  geom_point(mapping = aes(x = GHG_GR, y = Inflation_GR)) +
  geom_smooth(mapping = aes(x = GHG_GR, y = Inflation_GR), method = "lm") +
  labs(
    title = "Scatterplot of Greenhouse gas emission vs inflation rate",
    x = "Greenhouse gas emission",
    y = "Inflation rate")
```

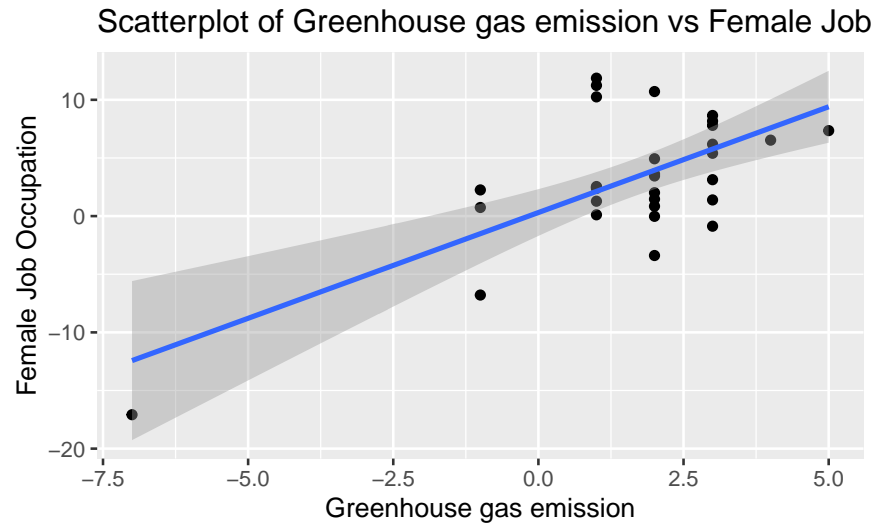
'geom_smooth()' using formula = 'y ~ x'



```
Group_dataset %>%
  ggplot() +
  geom_point(mapping = aes(y = GHG_GR, x = FJO_GR)) +
```

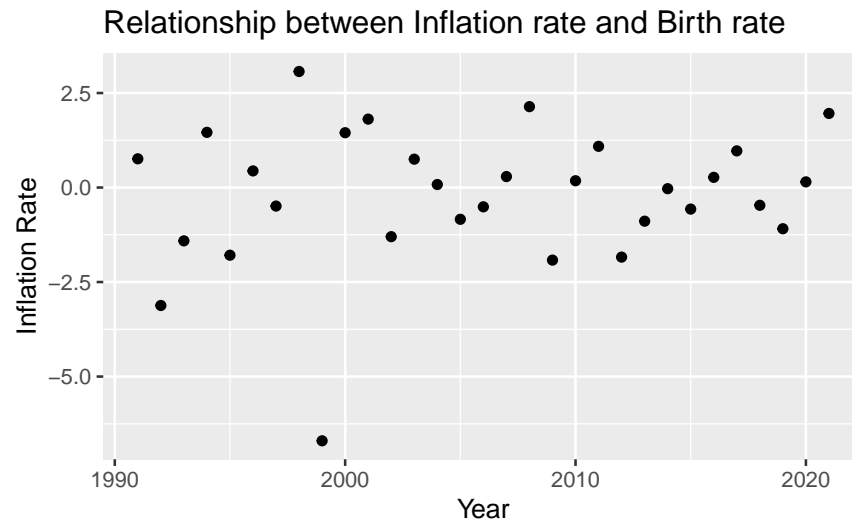
```
geom_smooth(mapping=aes(y = GHG_GR, x = FJO_GR), method="lm")+
  labs(
    title = "Scatterplot of Greenhouse gas emission vs Female Job Occupation",
    x = "Greenhouse gas emission",
    y = "Female Job Occupation")
```

'geom_smooth()' using formula = 'y ~ x'



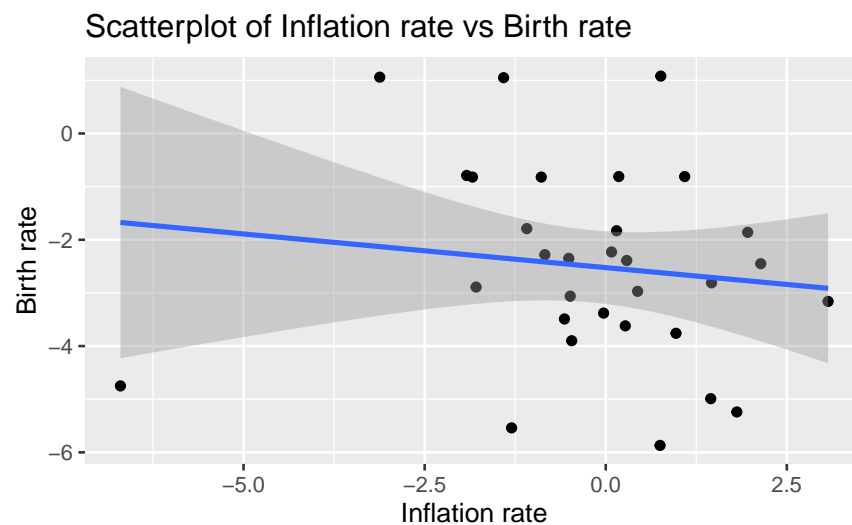
#Variation and Covariation - Econ_ft

```
Econ_ft %>%
  ggplot() +
  geom_point(mapping = aes(y = Inflation_GR, x= Year))+
  labs(title = "Relationship between Inflation rate and Birth rate",
    y = "Inflation Rate",
    x = "Year")
```



```
Econ_ft %>%
  ggplot() +
  geom_point(mapping=aes(x = Inflation_GR, y = Birth_GR))+
  geom_smooth(mapping=aes(x = Inflation_GR, y = Birth_GR), method='lm')+
  labs(
    title= 'Scatterplot of Inflation rate vs Birth rate',
    x= 'Inflation rate',
    y= "Birth rate")
```

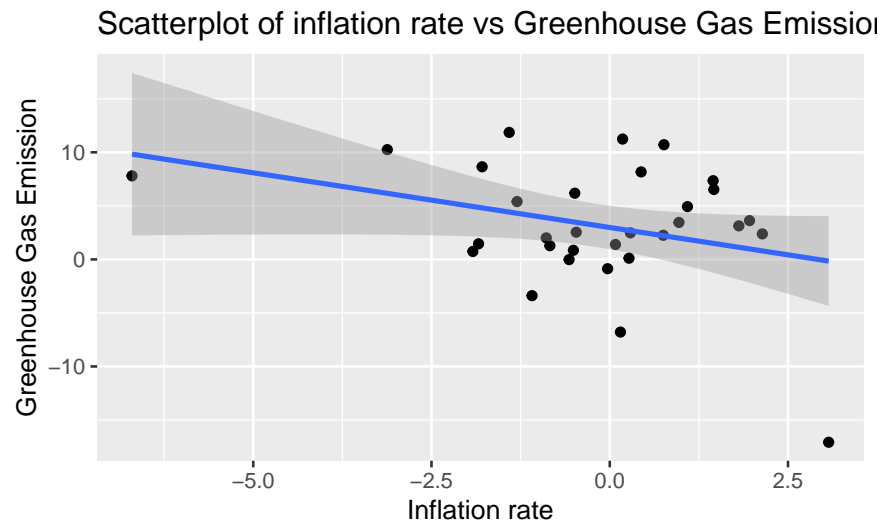
'geom_smooth()' using formula = 'y ~ x'



```
Group_dataset %>%
  ggplot() +
  geom_point(mapping= aes( x =Inflation_GR, y = GHG_GR))+
```

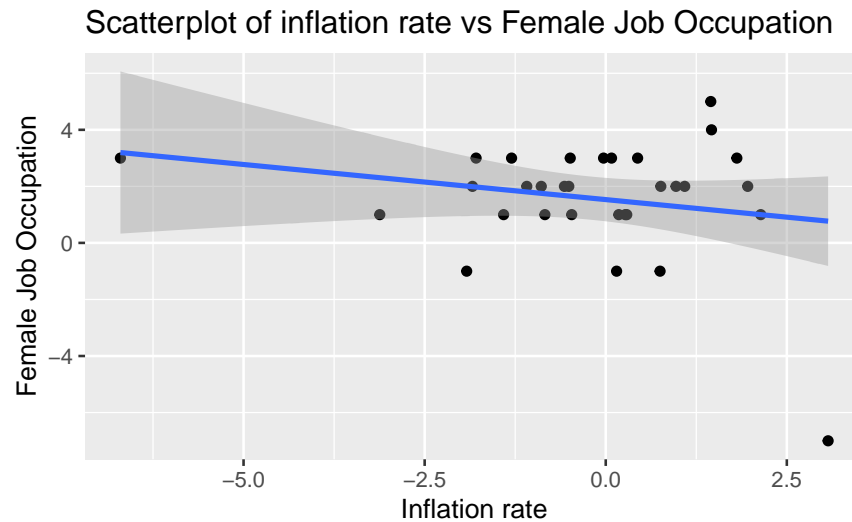
```
geom_smooth(mapping=aes(x = Inflation_GR, y = GHG_GR),
method= 'lm')+
  labs(
    title= 'Scatterplot of inflation rate vs Greenhouse Gas Emission',
    x= 'Inflation rate',
    y= 'Greenhouse Gas Emission')
```

'geom_smooth()' using formula = 'y ~ x'



```
Group_dataset %>%
  ggplot() +
  geom_point(mapping = aes(x= Inflation_GR, y= FJO_GR))+
  geom_smooth(mapping=aes(x= Inflation_GR, y= FJO_GR),
method= 'lm')+
  labs(
    title= 'Scatterplot of inflation rate vs Female Job Occupation',
    x= 'Inflation rate',
    y= 'Female Job Occupation')
```

'geom_smooth()' using formula = 'y ~ x'



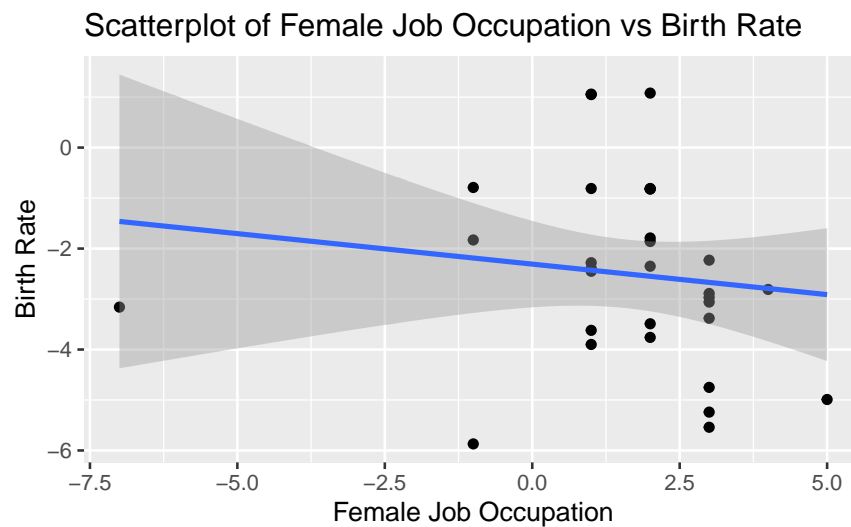
```
#Scatterplot
Soci_ft %>%
  ggplot() +
  geom_point(mapping = aes(y = FJO_GR, x = Year)) +
  labs(title = "Relationship between Female Job Occupation and Birth Rate", y = "Female Job Occupation")
```



```
# Scatterplot of Female Job Occupation vs Birth Rate
Soci_ft %>%
  ggplot() +
  geom_point(mapping = aes(x= FJO_GR, y= Birth_GR))+
  geom_smooth(mapping=aes(x = FJO_GR, y = Birth_GR), method="lm")+
  labs(
    title="Scatterplot of Female Job Occupation vs Birth Rate",
    x= "Female Job Occupation",
```

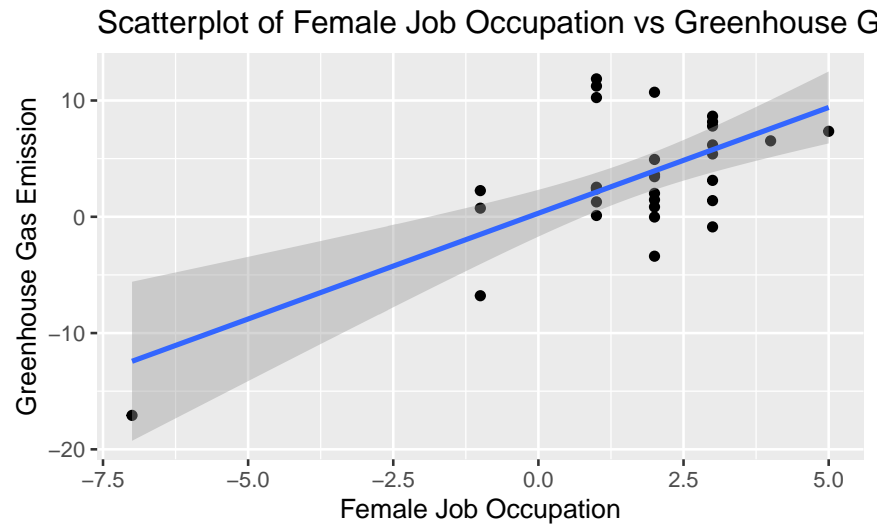
```
y= "Birth Rate"
)
```

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



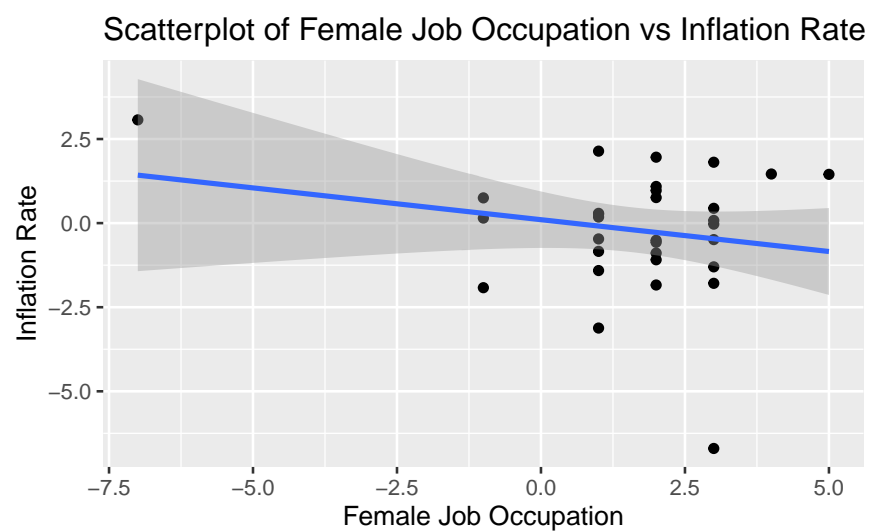
```
# Scatterplot of Female Job Occupation vs Greenhouse Gas Emission
Group_dataset %>%
  ggplot() +
  geom_point(mapping= aes (x= FJO_GR, y= GHG_GR)) +
  geom_smooth(mapping=aes (x= FJO_GR, y=GHG_GR), method="lm")+
  labs(
    title= "Scatterplot of Female Job Occupation vs Greenhouse Gas Emission",
    x= "Female Job Occupation",
    y= "Greenhouse Gas Emission"
  )
```

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

```
# Scatterplot of Female Job Occupation vs Inflation Rate
Group_dataset %>%
  ggplot() +
  geom_point(mapping= aes (x= FJO_GR, y=Inflation_GR)) +
  geom_smooth(mapping=aes (x= FJO_GR, y=Inflation_GR), method="lm")+
  labs(
    title= "Scatterplot of Female Job Occupation vs Inflation Rate",
    x= "Female Job Occupation",
    y= "Inflation Rate"
  )
```

'geom_smooth()' using formula = 'y ~ x'



```
# Model
Envo_ft_model <- lm(Birth_GR ~ GHG_GR, data = Envo_ft)
```

```
# Tidy model
Envo_ft_model %>%
  tidy()
```

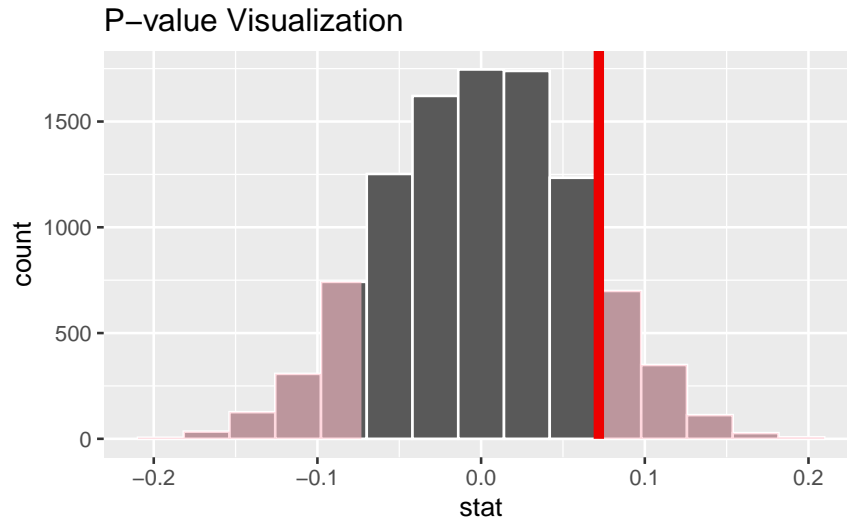
term	estimate	std.error	statistic	p.value
(Intercept)	-2.7278213	0.3752186	-7.269953	0.0000001
GHG_GR	0.0719243	0.0580149	1.239755	0.2250047

```
# Null distribution
Envo_null_distribution <- Envo_ft %>%
  specify(Birth_GR ~ GHG_GR) %>%
  hypothesize(null="independence") %>%
  generate(reps=10000, type="permute") %>%
  calculate(stat="slope")
```

```
# Observed stat
Observed_stat <- Envo_ft %>%
  specify(Birth_GR ~ GHG_GR) %>%
  calculate(stat="slope")
```

```
# P-value
p_value <- Envo_null_distribution %>%
  get_p_value(obs_stat=Observed_stat, direction="both")
```

```
# P-value visualization
Envo_null_distribution %>%
  visualize() +
  shade_p_value(obs_stat=Observed_stat, direction= "both") +
  labs(title = "P-value Visualization")
```



```
# Model
Econ_ft_model <- lm(Birth_GR ~ Inflation_GR, data=Econ_ft)
```

```
# Tidy model
Econ_ft_model %>%
  tidy()
```

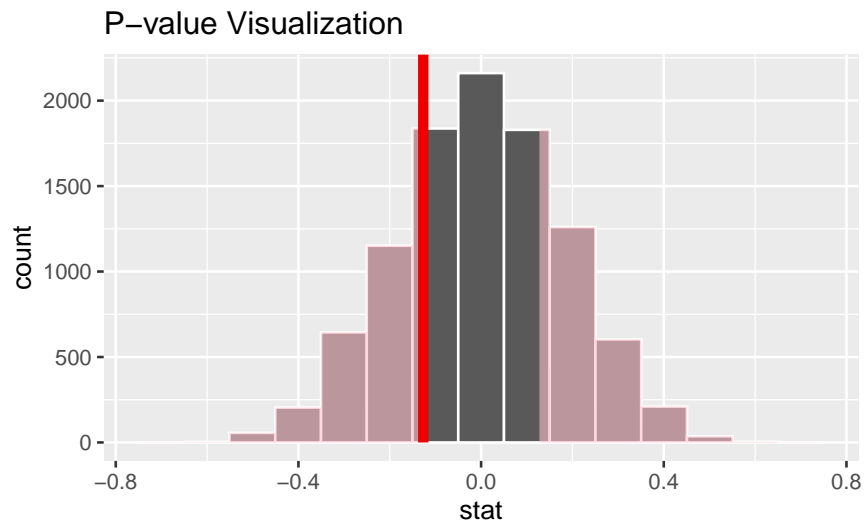
term	estimate	std.error	statistic	p.value
(Intercept)	-2.5239630	0.3345546	-7.5442476	0.0000000
Inflation_GR	-0.1266971	0.1850287	-0.6847431	0.4989426

```
# Null distribution
Econ_null_distribution <- Econ_ft %>%
  specify(Birth_GR ~ Inflation_GR) %>%
  hypothesize(null="independence") %>%
  generate(reps=10000, type="permute") %>%
  calculate(stat="slope")
```

```
# Observed stat
Observed_stat <- Econ_ft %>%
  specify(Birth_GR ~ Inflation_GR) %>%
  calculate(stat="slope")
```

```
# P-value
p_value <- Econ_null_distribution %>%
  get_p_value(obs_stat=Observed_stat, direction="both")
```

```
# P-value visualization
Econ_null_distribution %>%
  visualize() +
  shade_p_value(obs_stat=Observed_stat, direction= "both") +
  labs(title = "P-value Visualization")
```



#Modeling & Hypothesis test - Soci

```
# Model
Soci_ft_model <- lm(Birth_GR ~ FJO_GR, data = Soci_ft)
```

```
# Tidy model
Soci_ft_model %>%
  tidy()
```

term	estimate	std.error	statistic	p.value
(Intercept)	-2.3081339	0.4185330	-5.5148199	0.0000061
FJO_GR	-0.1207724	0.1611897	-0.7492564	0.4597376

```
# Null distribution
Soci_null_distribution <- Soci_ft %>%
  specify(Birth_GR ~ FJO_GR) %>%
  hypothesize(null="independence") %>%
  generate(reps=10000, type="permute") %>%
  calculate(stat="slope")
```

```
# Observed stat
Observed_stat <- Soci_ft %>%
  specify(Birth_GR ~ FJO_GR) %>%
  calculate(stat="slope")
```

```
# P-value
p_value <- Soci_null_distribution %>%
  get_p_value(obs_stat=Observed_stat, direction="both")
```

```
# P-value visualization
Soci_null_distribution %>%
  visualize() +
  shade_p_value(obs_stat=Observed_stat, direction="both") +
  labs(title = "P-value visualization")
```

