Final Project Writeup
[ECON140] Seamus Li
Aivan Durfee, Barton Zhuang, Christopher Adams, Josh Lim, Vincent Le
May 5, 2024

Replication of "*Every Day Is Earth Day: Evidence on the Long-Term Impact of Environmental Activism"*

1. **Background to Study:**

The paper *"Every Day Is Earth Day: Evidence on the Long-Term Impact of Environmental Activism"* by economists Daniel Hungerman and Vivek Moorthy studies the long term impacts of the first Earth Day (that occurred in 1970) on societal outlook affecting environmental perspectives. The first Earth Day (on April 22, 1970) encompassed multiple volunteer activities such as cleanups, protests, and rehabilitation efforts. The idea of Earth Day was to create a day where the environment was respected – whether it was through values, quality of efforts, or community health. The paper itself looks at how weather patterns (specifically negative weather such as rain and snow) in specific areas around the world on Earth Day may have affected Earth day participation, which in turn may have weakened environmental activism or health indicators in that area. This paper previously builds upon several other papers that explored the short term effects of activism and government policies on environmental factors but that did not explore long term impacts of volunteer driven actions on the environment.

---

2. **Summary of Study:**

Throughout our group's analysis of the research conducted by Hungerman and Moorthy, the paper uses multiple pieces of data (which we will detail later), to provide a comprehensive examination of the impact of environmental activism through Earth Day on long term public attitudes, health outcomes and environmental policy. The study uses a natural experiment, weather variation on April 22, 1970, as an instrumental variable to analyze how participation in Earth Day activities may have affected the societal outlook in several different areas. The economists hypothesize that negative weather on Earth Day led to reduced participation in environmental activities (which is quite logical), and in turn could have affected the development of positive environmental attitudes and initiatives in the specific communities. Data was collected through various public databases such as the General Social Survey (GSS) which included data ranging from 1977 to 1993. The GSS explored attitudes towards environmental spending in places with varying weather conditions on Earth Day, which aligned with the paper's purpose. The survey mainly detailed that poor weather on Earth Day exhibited significantly less support for environmental spending (in the subsequent 10 to 20 years), especially among the younger population that included children or teenagers at the time of Earth Day.

While the paper analyzed surveys like the GSS, it also utilized statistical data such as air pollution indicators, carbon monoxide measurements, and local pollutant levels to assess the long term effects of Earth Day's weather on environmental quality. This section of the paper was quite number based and was quite hard for our group to interpret at first. However, findings indicated that regions with negative weather on Earth Day experience much higher levels of carbon monoxide in the atmosphere over the subsequent two decades, reinforcing the paper's original hypothesis (the impact of Earth Day participation may have some effect on local environmental quality). Health consequences were also looked at, specifically the congenital abnormalities in newborns – a 1 standard deviation increase in precipitation on Earth Day was associated with a (small) but statistically significant increase in the incidence of congenital abnormalities in children born in the subsequent 20 years (this was evident among low socioeconomic family status). Hungerman and Moorthy also used several robustness checks that included placebo tests for rainfall data, statistical controls, and fixed effects to account for other factors. The results of the models and figures formulated through data regression processes only strengthened the validity of their original hypothesis.

---

3. Regression Analysis Used in Study:

For this study, research was conducted in 2022 and a Two-Stage Least Squares (2SLS) regression model was utilized to study the long-term effects of the first Earth Day in 1970. By doing so, Hungerman and Moorthy aimed to isolate the impact of environmental activism from many confounding factors, notably that of regional weather variations.

Amount of precipitation on Earth Day is the key independent variable of interest for the analysis, and is used as an instrumental variable to illustrate the level of participation in Earth Day activities. Since the authors conducted the study on the premise that poor weather would likely discourage outdoor civic participation, this variable was highlighted as a means of understanding the long-term influence of the Earth Day events on community attitudes and behaviors regarding the environment.

$$(1) \qquad y_{ct} = \alpha + r_c\phi + X_{ct}\beta + \Gamma_c\delta + \Phi_t\lambda + e_{ct},$$

$$(2) \qquad y_{ict} = \alpha + r_c\phi + X_{ict}\zeta + X_{ct}\beta + \Gamma_c\delta + \Phi_y\lambda + e_{ict},$$

Included above are two equations representing the 2SLS regression used in this study, which both have an additional error term at the end. The use of geographic and time-fixed effects is necessary to control for potentially unobservable factors that could bias the results. As an example, unique geographical effects of a given place/region may see inherent environmental or cultural differences that may lead to varied levels of community responsiveness to Earth Day.

Furthermore, the authors conducted robustness checks in order to validate their findings. These checks included testing different specifications and utilizing alternative measures of precipitation to ensure results were not driven by any particular model choices and were clean of outliers. Communities that experienced unusually high or low amounts of rainfall were compared and “standardized” to their typical climate patterns, enhancing the credibility of their findings by focusing on the usual weather patterns of a place (as this is what drives the observed outcome) rather than solely the level of rainfall. By analyzing this rigorous statistical framework that Hungerman and Moorthy implemented, we can begin to better understand how environmental/activist events can influence long-term public perceptions and health.

Ultimately, our group believes that this study proved the ability of environmental activism as a mechanism for influencing community perspectives on environmental expenditure, local air quality, and the well-being of subsequent generations; Hungerman and Moorthy's study heavily emphasizes the crucial role of *collective* and civic involvement in tackling environmental challenges, showing that initiatives such as Earth Day can yield significant benefits for both the societies of today and the future.

---

4. Conclusion:

Our replication of Hungerman and Moorthy's study reaffirmed the robustness of their research methodology. Using a Two-Stage Least Squares (2SLS) regression model, we tried our best to match their results. One important finding was that upon isolating for weather conditions, we were able to confirm that adverse weather did, in fact, cause a negative impact on long-term community support for environmental initiatives. Overall, this supports the idea that weather patterns can impact participation in significant environmental events, which in turn can shape public attitudes and behaviors toward environmental issues and informed policy making. Through our replication of Hungerman and Moorthy's study, we gained valuable insights into the enduring impacts of the first Earth Day as well as the necessity for methodological rigor in the research and data collection processes (for replicability particularly).

---

5. Replication Results:

The replication process was quite challenging for our group. The dataset included by the paper's authors contained a lot of individual Stata files that we had to refine. Additionally, we had to learn the process of how the data was filtered from the appendix and then translate that into our code. We found converting Stata code to R code challenging, as this entire study was done in Stata. Additionally, there was minimal discussion (in the appendix) of the exact replication process that the authors used to generate tables and charts, so much of our process was spent playing with the data on our own. Lastly, To gain a comprehensive understanding of the datasets, we referenced the data glossaries provided in the readme file for each dataset. By consulting these glossaries, we aimed to decipher the meanings of the unique column names, enabling us to identify which data columns were utilized in our study. This approach facilitated our efforts to replicate the study accurately and ensured that we selected the appropriate data for analysis.

We chose to replicate Figure 4, as this meant that we had to produce a regression summary of the 2SLS (with an instrument variable) described above and then chart this for each of the days in the month to match the original chart provided by the authors. Figure 4 displays a day-by-day analysis of the effects of deviations from historical average precipitation in April 1970 on carbon monoxide levels measured from 1970 to 1988 (on the graph, the data is in parts per million). This represents the relationship between rainfall from April 17th to the 28th, 1970 and carbon monoxide levels in the air to see if there was a possible causal relationship in this aspect. The x-axis of Figure 4 represents days in April 1970 (surrounding the first Earth day), and the y-axis quantifies the impact of precipitation deviations on carbon monoxide levels. As a group, we believe that the graph is designed to show how unusual weather on a specific Earth Day might correlate with long-term environmental quality, indicated by changes in carbon monoxide concentration in the atmosphere.

In parallel to Figure 4, we also replicated Figure 5 by producing another 2SLS regression summary of congenital malformations from 1980 to 1988 on a set of covariates from rainfall during April 17th to the 28th, 1970. The day-by-day results highlight the association between adverse weather conditions, specifically rainy weather, during Earth Day and an increased incidence of congenital malformations occurring 10 to 20 years later. As per the study, the coefficients are multiplied by a magnitude of 100, indicating that a 100th tenth of a millimeter increase in rain corresponds to a 0.3 percent rise in the probability of a child being born with a congenital malformation. Furthermore, one standard deviation increase in rain increases the percentage of children born with congenital malformation by an estimated 0.1 standard deviations. As a group, this model illustrates the long-term benefits of environmental initiatives on newborn health, highlighting continuous advantages irrespective of fetal origins arguments. This is especially significant because the people studied were not even born on the first Earth Day (April 22nd, 1970).

We have also chosen to replicate Table 3 in the study additionally. Table 3 examines the relationship between rainfall on Earth Day, April 22, 1970, and subsequent environmental outcomes. It specifically examines how deviations in rainfall correlate with carbon monoxide (CO) pollution levels and congenital abnormalities. Our group believes that this reveals that increased rainfall on Earth Day is associated with higher CO levels over the next 20 years and a greater occurrence of congenital abnormalities in children born 10-20 years later. This relationship is particularly pronounced among low-socioeconomic-status mothers, suggesting that adverse weather conditions during Earth Day had long-term impacts on both environmental quality and public health, especially in vulnerable communities. We found this replication to be especially tricky due to several unforeseen factors. Firstly, the paper lacks any coherent description of the data-cleaning processes that they used, making it extra difficult to replicate exactly. We had to create our own cleaning process that was less comprehensive than the original papers. Additionally, the study's heavy reliance on R made it difficult to directly translate any of the data or DO files into usable R code. Overall, the absence of explicit instructions for merging multiple datasets and also handling missing data was the trickiest part of this replication. These obstacles highlight just some of the barriers that we had to overcome to try and achieve a perfect replication. We learned that subtle nuances in the management of the data and also the statistical analyses that we used greatly affect the ability to replicate this table. Had the methodology been more transparent from the original authors, the replication would have been almost perfect.

See the replications below:

Figure 4 – ORIGINAL:

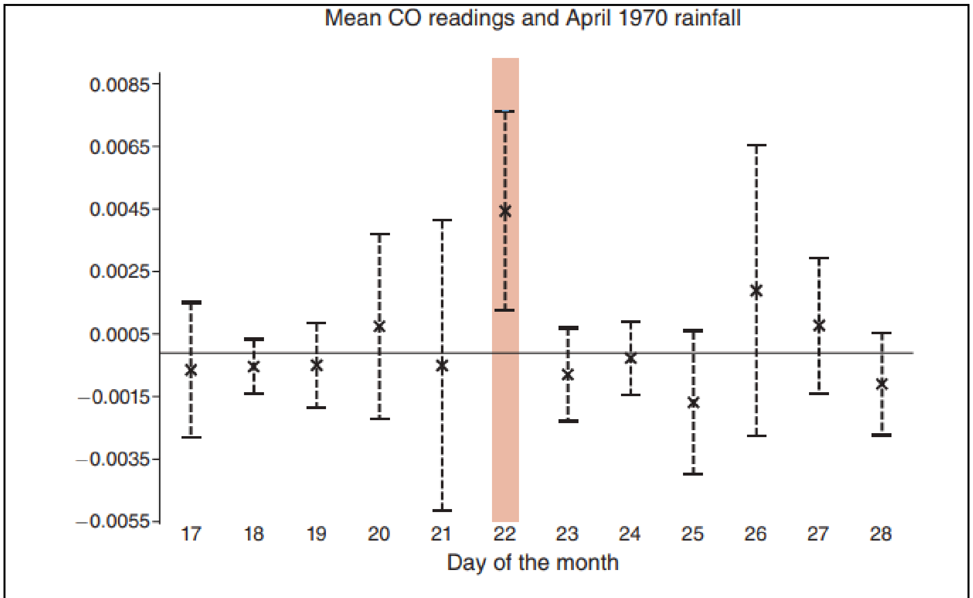

FIGURE 4. EARTH DAY AND AIR POLLUTION

Figure 4 – REPLICATED:

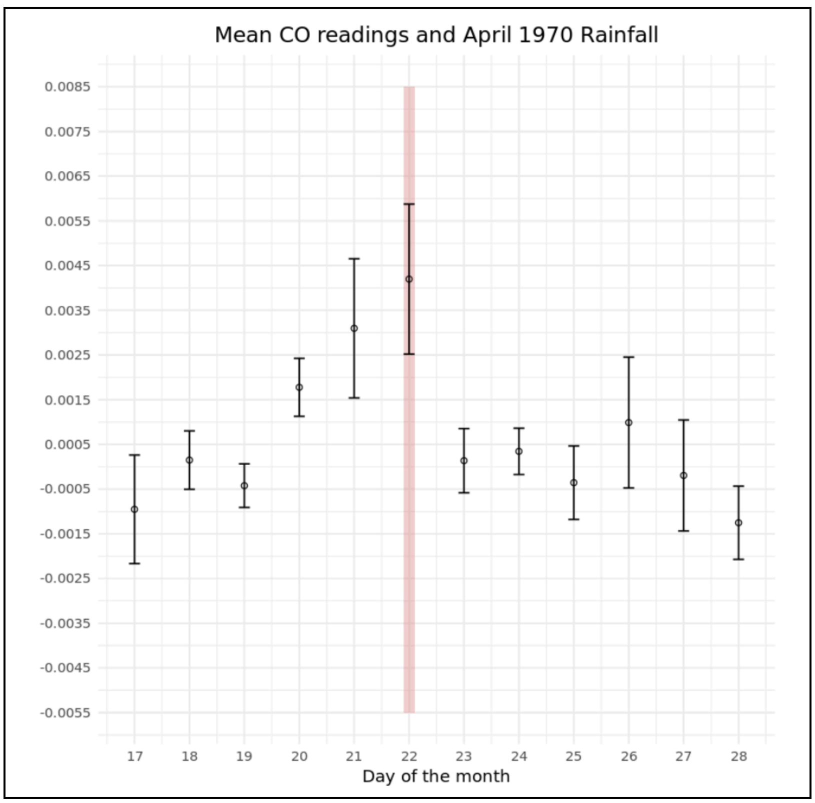

Figure 5 – ORIGINAL:

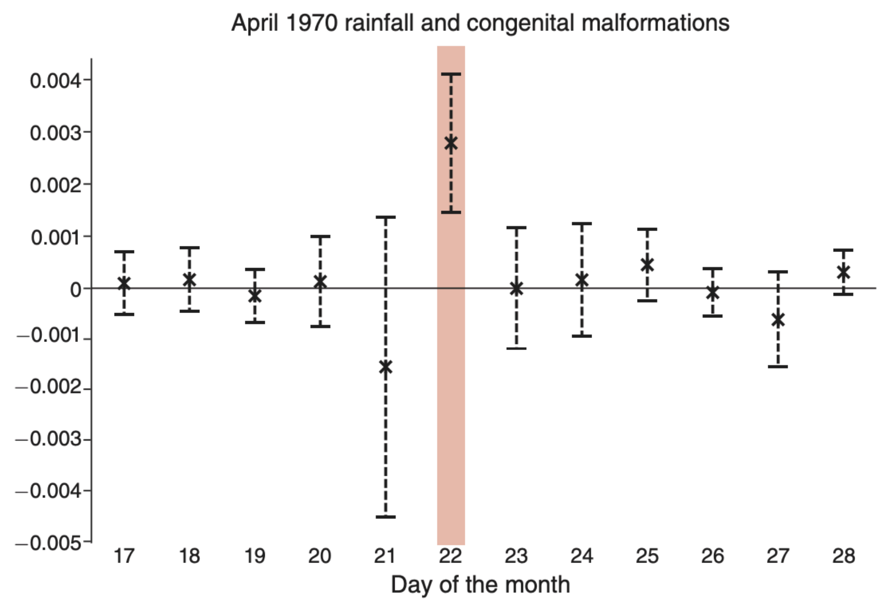

FIGURE 5. EARTH DAY AND INFANT HEALTH

Figure 5 – REPLICATED:

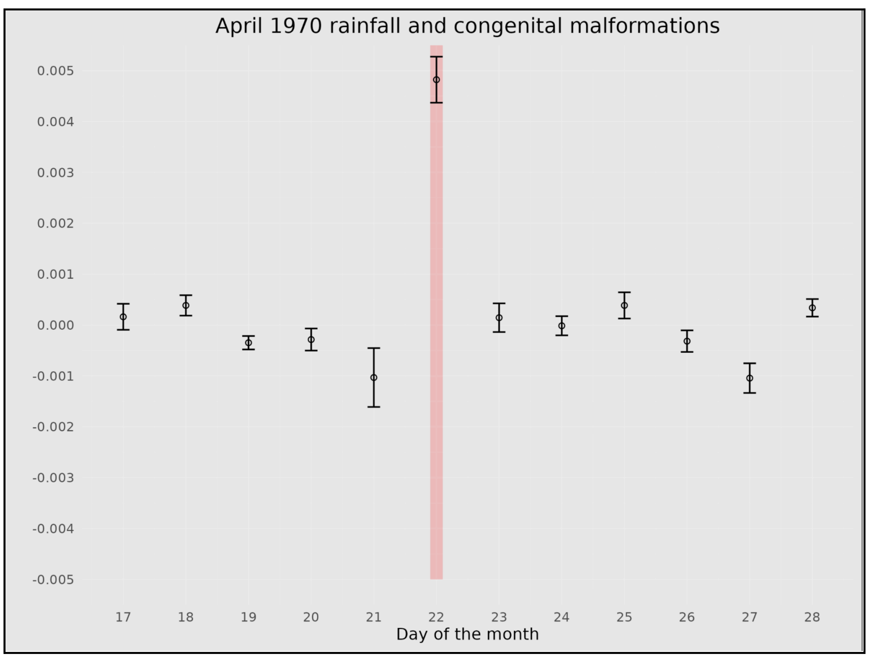

Table 3 – ORIGINAL:

	Carbon monoxide (1)	Congenital abnormalities† All (2)	High-SES births (3)	Low-SES births (4)
Rain on Earth Day	0.00360 (0.00158)	0.00514 (0.0011)	0.00525 (0.0010)	0.00515 (0.0012)
Rain on Earth Day (levels)	0.00295 (0.00198)	0.00442 (0.0013)	0.00422 (0.0012)	0.00329 (0.0014)
Rain on Earth Day (extra controls)	0.00238 (0.00167)	0.00364 (0.0012)	0.0036 (0.0011)	0.00474 (0.00118)
Rain on Earth Day (weighted, extra controls)	0.0031 (0.00161)	0.00475 (0.0019)	0.00482 (0.0016)	0.00685 (0.00229)
Winzorized rainfall	0.00265 (0.00188)	0.00542 (0.0023)	0.00537 (0.0021)	0.00743 (0.00242)
Winzorized rainfall (weighted)	0.00344 (0.00193)	0.00628 (0.0027)	0.00679 (0.0024)	0.00934 (0.00308)

Table 3 – REPLICATED:

	Dependent variable: Carbon Monoxide (1)	Congenital Abnormalities (2)
Rain on Earth Day	-0.0005* (0.0002)	-0.00000** (0.00000)
Constant	1.375*** (0.016)	0.013*** (0.0001)
Observations	2,209	9,323
R2	0.001	0.0003
Adjusted R2	0.001	0.0002
Residual Std. Error	0.683 (df = 2207)	0.013 (df = 9321)
F Statistic	2.483 (df = 1; 2207)	2.558 (df = 1; 9321)
Note:		*p<0.1; **p<0.05; ***p<0.01

6. Citations:

Hungerman, Daniel, and Vivek Moorthy. 2023. "Every Day Is Earth Day: Evidence on the Long-Term Impact of Environmental Activism." American Economic Journal: Applied Economics 15 (1): 230–58. https://doi.org/10.1257/app.20210045.

Code Appendix:

Figure 4 Code:

```
# Loading libraries
library('haven')
library(ggplot2)
library(dplyr)

# Reading in data sets
AQI <- read_dta('Rep_Data/AQI Polluants from Monitors.dta')
april_rainfall <- read_dta("Rep_Data/rain_many_days.dta")

# Grabbing columns to pick targeted data
AQI_cols <- names(AQI)
AQI_selected <- AQI[, c("year", "fips", "co")]

# Filtering data on range 1970 - 1988 for year column
AQI_selected_1970 <- AQI_selected[AQI_selected$year >= 1970 & AQI_selected$year <= 1988,]
# Omitting NA rows
AQI_final <- na.omit(AQI_selected_1970)

# Grouping data to aggregate mean carbon monoxide
AQI_final_mean <- AQI_final %>% group_by(fips, year) %>% summarize(mean_co = mean(co, na.rm =
TRUE))

# Extracting specified rainfall days (days 17 to 28 of April 1970)
d_april_rainfall_selected <- april_rainfall[, c("fips", "dPRCP_day1770", "dPRCP_day1870",
"dPRCP_day1970", "dPRCP_day2070", "dPRCP_day2170", "dPRCP_day2270", "dPRCP_day2370",
"dPRCP_day2470", "dPRCP_day2570", "dPRCP_day2670", "dPRCP_day2770", "dPRCP_day2870")]

# Omitting NA rows
d_april_rainfall_filtered <- na.omit(d_april_rainfall_selected)

# Merging two data sets together on fibs
merged_df <- inner_join(d_april_rainfall_filtered, AQI_final_mean, by = c("fips"))

# Running regression model
model <- lm(mean_co ~ dPRCP_day1770 + dPRCP_day1870 + dPRCP_day1970 + dPRCP_day2070 +
dPRCP_day2170 + dPRCP_day2270 + dPRCP_day2370 + dPRCP_day2470 + dPRCP_day2570 +
dPRCP_day2670 + dPRCP_day2770 + dPRCP_day2870, data = merged_df)
summary(model)
```

```
# Preparing model data for graphing
coef_df <- coef(model)
coef_df <- coef_df[2:13]
coef_names <- c(17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28)
conf_int <- confint(model, level = 0.95)
ci_low <- conf_int[, 1]
ci_low <- ci_low[2:13]
ci_high <- conf_int[, 2]
ci_high <- ci_high[2:13]
summary_model <- coef(summary(model))
std <- summary_model[, "Std. Error"]
std <- std[2:13]
highlight <- data.frame(
day = 22, # Day to highlight
ymin = -0.0055, # Lower limit for y-axis
ymax = 0.0085 # Upper limit for y-axis)
df <- data.frame(coef_df, coef_names, ci_low, ci_high)

# Graphing model
plot <- ggplot() + geom_rect(data = highlight, aes(xmin = day - 0.1, xmax = day + 0.1, ymin =
ymin, ymax = ymax), fill = "lightcoral", alpha = 0.4) + geom_point(data = df, aes(x = coef_names, y =
coef_df), shape = 1) +   geom_errorbar(data = df, aes(x = coef_names, y = coef_df, ymin = ci_low, ymax
= ci_high), width = 0.2) +   labs(title = "Mean CO readings and April 1970 Rainfall", x = "Day of the
month", y = "") + coord_cartesian(ylim = c(-0.0055, 0.0085)) + scale_y_continuous(breaks = seq(-0.0055,
0.0085, by = 0.001)) + scale_x_continuous(breaks = seq(17, 28, by = 1)) + theme_minimal() +
theme(plot.title = element_text(hjust = 0.5, family = "roboto", size = 14))
ggsave("plot.png", plot, width = 8, height = 6)
```

Figure 5 Code:

```
# install.packages('haven')
library(haven)

# install.packages('ggplot2')
library(ggplot2)

# install.packages('dplyr')
library(dplyr)
```

```
# Read data
april_rainfall <- read_dta("ECON-140-SP24/rain_many_days.dta")
bweight_data <- read_dta("ECON-140-SP24/bweight_79_88.dta")
bweight_data <- bweight_data %>%
  filter(year >= 1980)

# Inspect the structure of the data to find the necessary variables
str(april_rainfall)
str(bweight_data)

# Find the # of births in the table
total_value <- sum(bweight_data$nbirths)

# Filter data for necessary columns
bweight_filtered <- bweight_data %>%
  select(countyfip, congen, year) %>%
  group_by(countyfip) %>%
  summarize(
    mean_congen = mean(congen),
    sum_congen = sum(congen),
    congen = congen
  )

april_filtered <- april_rainfall %>%
select(fips, dPRCP_day1770, dPRCP_day1870, dPRCP_day1970, dPRCP_day2070, dPRCP_day2170 ,
dPRCP_day2270, dPRCP_day2370, dPRCP_day2470, dPRCP_day2570, dPRCP_day2670,
dPRCP_day2770, dPRCP_day2870)

april_filtered <- na.omit(april_filtered)

# Merge the two tables
merged_df <- inner_join(april_filtered, bweight_filtered, by = c("fips" = "countyfip"))

# Model on covariates from 4/17/70 to 4/28/70 for congen
model <- lm(congen ~ dPRCP_day1770 + dPRCP_day1870 + dPRCP_day1970 + dPRCP_day2070 +
dPRCP_day2170  + dPRCP_day2270 + dPRCP_day2370 + dPRCP_day2470 + dPRCP_day2570 +
dPRCP_day2670 + dPRCP_day2770 + dPRCP_day2870, data = merged_df)
summary(model)

# Get coefficients, feature names, and confidence intervals (Multiplied by 100 for readability)
summary_model <- coef(summary(model))
```

```
print(summary_model)

coef_df <- coef(model) * 100
coef_df <- coef_df[2:13]

coef_names <- c(17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28)

conf_int <- confint(model, level = 0.95) * 100
ci_low <- conf_int[, 1]
ci_low <- ci_low[2:13]
ci_high <- conf_int[, 2]
ci_high <- ci_high[2:13]

std <- summary_model[, "Std. Error"] * 100
std <- std[2:13]

# Define highlight rectangle in the figure
highlight <- data.frame(
  day = 22, # Day to highlight
  ymin = -0.005, # Lower limit for y-axis
  ymax = 0.0055 # Upper limit for y-axis
)

# Create df for plot
df <- data.frame(coef_df, coef_names, ci_low, ci_high)
print(df)

plot <- ggplot() + geom_rect(data = highlight, aes(xmin = day - 0.1, xmax = day + 0.1, ymin = ymin,
ymax = ymax), fill = "lightcoral", alpha = 0.4) + geom_point(data = df, aes(x = coef_names, y = coef_df),
shape = 1) + geom_errorbar(data = df, aes(x = coef_names, y = coef_df, ymin = ci_low, ymax = ci_high),
width = 0.2) +  labs(title = "April 1970 rainfall and congenital malformations", x = "Day of the month", y
= "") +  coord_cartesian(ylim = c(-0.005, 0.005)) + scale_y_continuous(breaks = seq(-0.005, 0.005, by =
0.001)) + scale_x_continuous(breaks = seq(17, 28, by = 1)) +  theme_minimal() + theme(plot.title =
element_text(hjust = 0.5, family = "roboto", size = 14))
ggsave("plot.png", plot, width = 8, height = 6)
```