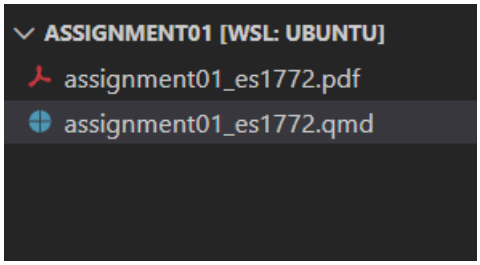
PPOL 6819 | Assignment 01

Ethan Sager

2026-01-20

0.1 Exercise 01

Below you can see an image of the directory for the assignment. Since I am working in vs code, there will be no .Rproj file shown.



0.2 Exercise 02

```
# 1) Load libraries
library(tidyverse)

# 2) Load data
energy <- read_csv(
    "https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2023/2023-06-
)

# 3) Glimpse data
glimpse(energy)
```

0.3 Exercise 03

```
# 1) count the number of obs per country

energy %>%
    count(country)

# 2) count the number of missing values for each variable using map_dbl

energy %>%
    map_dbl(~ sum(is.na(.)))
```

0.4 Exercise 04

```
# 1) Create a new data frame with only netherlands and france

countries <- c("Netherlands", "France")

energy_nd_fr <- energy %>%
    filter(country %in% countries)

# 2) divide the wind_cons_change_pct by 100 and call it wind_cons_change_prop
energy_nd_fr <- energy_nd_fr %>%
    mutate(wind_cons_change_prop = wind_cons_change_pct / 100)

# 3) group by country and calculate the year-to-year change

energy_nd_fr <- energy_nd_fr %>%
    group_by(country) %>%
    arrange(year) %>%
    mutate(
        wind_consumption_change = wind_consumption - lag(wind_consumption)
    ) %>%
    ungroup()

# 4) calculate the median wind consumption change pct and prop for each country since 2000

energy_nd_fr %>%
    filter(year >= 2000) %>%
```

```
    group_by(country) %>%
    summarize(
        median_wind_cons_change = median(wind_consumption_change, na.rm = TRUE),
        median_wind_cons_change_pct = median(
            wind_cons_change_pct,
            na.rm = TRUE
        ),
        median_wind_cons_change_prop = median(
            wind_cons_change_prop,
            na.rm = TRUE
        )
    )
```

0.5 Exercise 05

```
# 1) create energy us
energy_us <- energy %>%
    filter(country == "United States" & year > 1988)

# 2) fit a simple linear model
model <- lm(wind_electricity ~ year, data = energy_us)

# 3) summarize the model
summary(model)

# 4) extract r2
# Load broom (though note we can just take the base object and get r2 as well)
library(broom)
model_summary <- glance(model)
r_squared <- model_summary$r.squared
r_squared

# 5) interpret the coef for year
# We can see a postive realtionship between year and
# wind electricity generation in the US. The coefficient for year indicates that,
# on average, wind electricity generation increases
# by approximately 1,500 GWh each year this is highly signifigant at the < 1%

# 6) Given that the p-val is super tiny we know that the
# coef is signifigant and the odds of this being a
```

```
# fluke in the data distribution is extremely small.

# 7) The r2 of .7892 tells us that around 79% of the
# varaiation between obseravation is expalined by what year it is.
# Which makes alot of since as the cost of wind has gone down the capacity has gone up.
```

0.6 Exercise 06

Below you an see a scatter plot of wind electricity generation in the US over time with the linear model fit shown in red.

```
ggplot(energy_us, aes(x = year, y = wind_electricity)) +
    geom_point() +
    geom_smooth(method = "lm", col = "blue") +
    labs(
        title = "Wind Electricity Generation in the US Over Time",
        x = "Year",
        y = "Wind Electricity (GWh)"
    )
```