

Calculating ratios with R

ToDo

1. Load data
2. Preprocess and merge

1. Load data

```
library(tidyverse)
library(here)

## here() starts at /home/jules/proj/capex_green

ratio <- readxl::read_excel(here("data", "investments_lowcarbon+GS+JB.xlsx")) %>%
  select(company, year, `capex_low-carbon_as_pct_calc`) %>%
  mutate(low_carbon_ratio = `capex_low-carbon_as_pct_calc` / 100) %>%
  select(-`capex_low-carbon_as_pct_calc`) %>%
  filter(!is.na(low_carbon_ratio))

## New names:
## * `Notes` -> `Notes...6`
## * `` -> `...7`
## * `` -> `...11`
## * `Notes` -> `Notes...15`
## * `` -> `...16`
## * `Notes` -> `Notes...20`

glimpse(ratio)

## Rows: 42
## Columns: 3
## $ company      <chr> "exxon", "exxon", "exxon", "exxon", "exxon", "exxon", ~
## $ year         <dbl> 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, ~
## $ low_carbon_ratio <dbl> 0.01478852, 0.01293996, 0.01196458, 0.01120448, 0.012~

posts <- read_csv(here("data", "1_labels_and_data.csv")) %>%
  mutate(year = year(published_at)) %>%
  select(channel_name, green, brown, misc, year)
glimpse(posts)

## Rows: 500
## Columns: 5
## $ channel_name <chr> "ExxonMobil", "ExxonMobil", "ExxonMobil", "ExxonMobil", "~
## $ green        <lgl> FALSE, FALSE, TRUE, TRUE, FALSE, TRUE, FALSE, FALSE, FALS~
## $ brown        <lgl> TRUE, FALSE, TRUE, FALSE, FALSE, TRUE, FALSE, FALSE, FALS~
## $ misc         <lgl> FALSE, TRUE, FALSE, FALSE, TRUE, FALSE, TRUE, TRUE, TRUE, ~
## $ year         <dbl> 2021, 2021, 2021, 2021, 2021, 2021, 2020, 2020, 2020, 202~

matching <- read_csv(here("data", "1_matching.csv"))
glimpse(matching)
```

```
## Rows: 7
## Columns: 2
## $ channel <chr> "ExxonMobil", "ExxonMobil LNG", "ExxonMobil Nat Gas", "ExxonMo~
## $ company <chr> "exxon", "exxon", "exxon", "exxon", "exxon", "exxon", "exxon"
```

2. Preprocess and merge

```
aggregated_posts <- posts %>%
  left_join(matching, by = c("channel_name" = "channel")) %>%
  group_by(company, year) %>%
  summarize(green_ratio = sum(green) / sum(green | brown), .groups = "drop")
glimpse(aggregated_posts)
```

```
## Rows: 8
## Columns: 3
## $ company      <chr> "exxon", "exxon", "exxon", "exxon", "exxon", "exxon", "exx~
## $ year         <dbl> 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024
## $ green_ratio  <dbl> 0.0000000, 0.6923077, 0.4666667, 0.6000000, 0.4358974, 0.5~
```

```
ratios <- aggregated_posts %>%
  left_join(ratio, by = c("company", "year")) %>%
  mutate(posts_to_capex = green_ratio / low_carbon_ratio,
         posts_to_capex_normalized = (green_ratio - low_carbon_ratio) /
           (green_ratio + low_carbon_ratio))
glimpse(ratios)
```

```
## Rows: 8
## Columns: 6
## $ company      <chr> "exxon", "exxon", "exxon", "exxon", "exxon", "~
## $ year         <dbl> 2017, 2018, 2019, 2020, 2021, 2022, 2023, 20~
## $ green_ratio  <dbl> 0.0000000, 0.6923077, 0.4666667, 0.6000000, ~
## $ low_carbon_ratio <dbl> 0.02061431, 0.01838573, 0.01531159, 0.022251~
## $ posts_to_capex <dbl> 0.000000, 37.654615, 30.478000, 26.964000, 1~
## $ posts_to_capex_normalized <dbl> -1.0000000, 0.9482597, 0.9364636, 0.9284795,~
write_csv(ratios, here("data", "1_ratios.csv"))
```