Data Assets and Intangible Capital in Canadian Productivity

Matthew

2025-08-16

Overview

This project looks at financial data from Statistics Canada (Table 33-10-0499-01), which tracks balance sheets, income statements, and financial ratios for industries across Canada. The data set spans several years and breaks down different types of assets, liabilities, and other financial components.

The goal here is to get a better picture of how the financial structure of Canadian industries has been changing over time. A big focus is on the difference between tangible assets (things like buildings, equipment, and land) and intangible assets (things like intellectual property, goodwill, and software). Tangible assets have always been important, but intangible assets are becoming a bigger part of the economy, and it's interesting to see how that shift shows up in the numbers.

In this notebook, I:

- Clean and organize the data so it's easier to work with.
- Build some time-series visualizations to see how key financial components change year by year.
- Compare trends in tangible vs. intangible assets to understand how their roles are evolving across industries.

Data Sources

Statistics Canada. Table 33-10-0499-01 Balance sheet, income statement and taxation statistics with selected financial ratios, by financial industries. DOI: https://doi.org/10.25318/3310049901-eng

This dataset provides annual financial information for Canadian financial industries, covering things like balance sheets, income statements, taxes, and key financial ratios. Most importantly it has separated values for intangible and tangible assets.

There is also data within the Statistics Canada Analytical Studies Branch Research Paper Series "Data, Intangible Capital and Economic Growth in Canada" by Rupert Allen, Wulong Gu and Ryan Macdonald Linked here. This data is gathered from the authors' tabulations.

```
readRenviron("~/.Renviron")
data_path <- Sys.getenv("project_path")

#Importing the Balance Sheet Data
new_raw <- read.csv(file.path(data_path, "3310049901_databaseLoadingData.csv"))
df <- new_raw

#glimpse(df)
```

Recoding, Renaming, and Removing. (The holy trinity of data manipulation.)

```
#Recoding our year as a Date. helps with time series analysis
df$Year <- as.Date(paste0(df$REF_DATE, "-01-01"))

# Renaming for conciseness
df <- df %>%
  rename(Industry = North.American.Industry.Classification.System..NAICS.,
         Balance_items = Balance.sheet.and.income.statement.components..selected.financial.ratios)

# Keeping only these variables
df <- df %>%
  dplyr::select(Year, GEO, Industry, Balance_items, SCALAR_FACTOR, VALUE, DECIMALS)


# Recoding as factors
df <- df %>%
  mutate(Balance_items = as.factor(Balance_items),
         Industry = as.factor(Industry))

#unique(df$Balance_items)
```

Results Gathered from Canadian Balance Sheets.

Some notes regarding these graphs.

- We use Fixed Assets (net value) as a proxy for tangible assets. Generally all fixed assets are tangible. but not all tangible assets are fixed. so we're not totally encompassing all available tangible assets in the Canadian economy from these balance sheets.
- Both our measures for Fixed assets (tangible), and Intangible assets are reported in Net Value. this means we subtract capital depreciation to arrive at these values.

From these results we've generated three graphs...

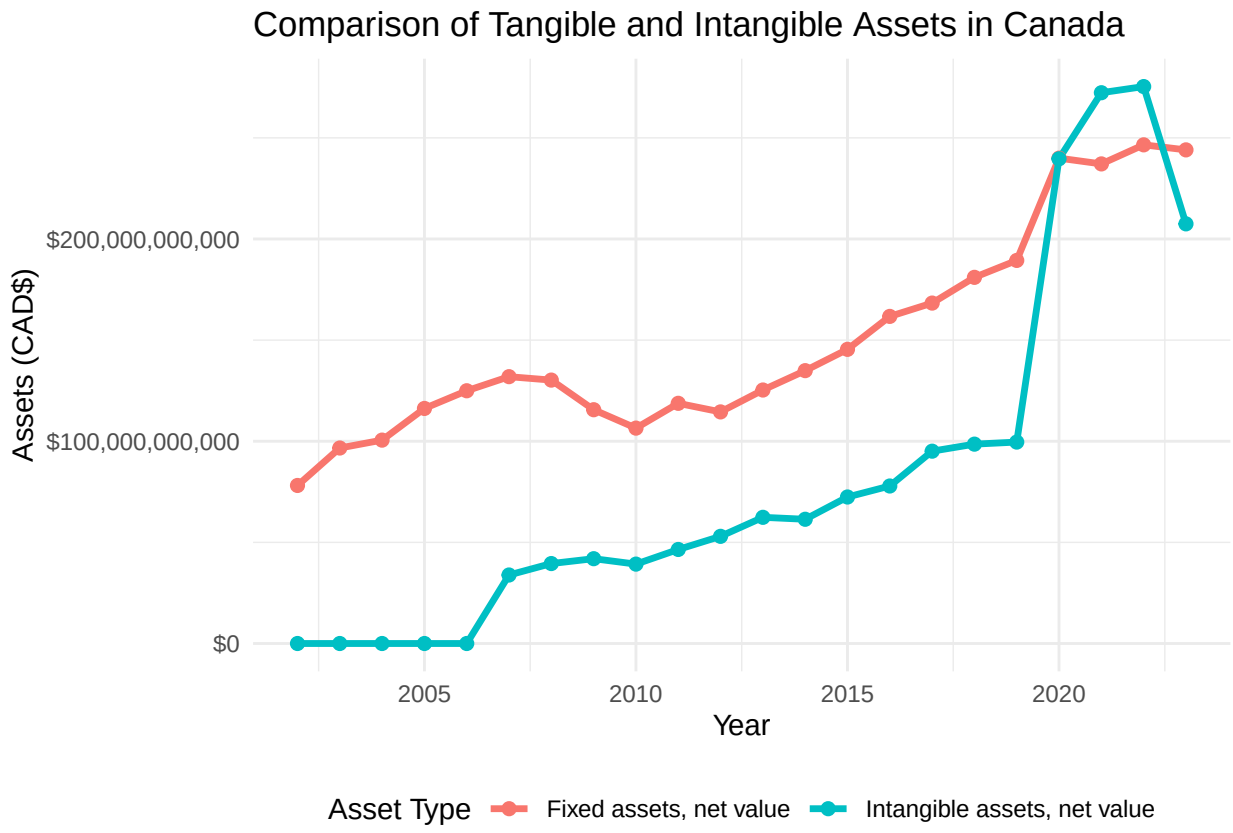
```
library(scales)

# Filter for tangible and intangible assets
df_assets <- df %>%
  filter(Balance_items %in% c("Intangible assets, net value", "Fixed assets, net value")) %>%
  mutate(
    VALUE_scaled = case_when(
      SCALAR_FACTOR == "millions" ~ VALUE * 1e6,
      SCALAR_FACTOR == "thousands" ~ VALUE * 1e3,
      TRUE ~ VALUE
    ) / (10^DECIMALS)
  )

# Summarize by year (Canada)
df_assets_canada <- df_assets %>%
  filter(GEO == "Canada") %>%
```

```r
  group_by(Year, Balance_items) %>%
  summarise(Total_Assets = sum(VALUE_scaled, na.rm = TRUE), .groups = "drop")

# Plot comparison
ggplot(df_assets_canada, aes(x = Year, y = Total_Assets, color = Balance_items)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  scale_y_continuous(labels = scales::dollar_format()) +
  labs(
    title = "Comparison of Tangible and Intangible Assets in Canada",
    x = "Year",
    y = "Assets (CAD$)",
    color = "Asset Type"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")
```



This graph shows a raw dollar amount of tangible and intangible capital.

Note:

- Consider the authors briefing on how to define intangible capital.
  - Case 1—Intangible assets currently classified as final consumption
  - Case 2—Intangible assets currently classified as intermediate inputs
  - Case 3—Intangible assets not currently included in national accounts

- In 2006 the nature of reporting intangible assets changed. which is why we have some non-0 values prior to then.

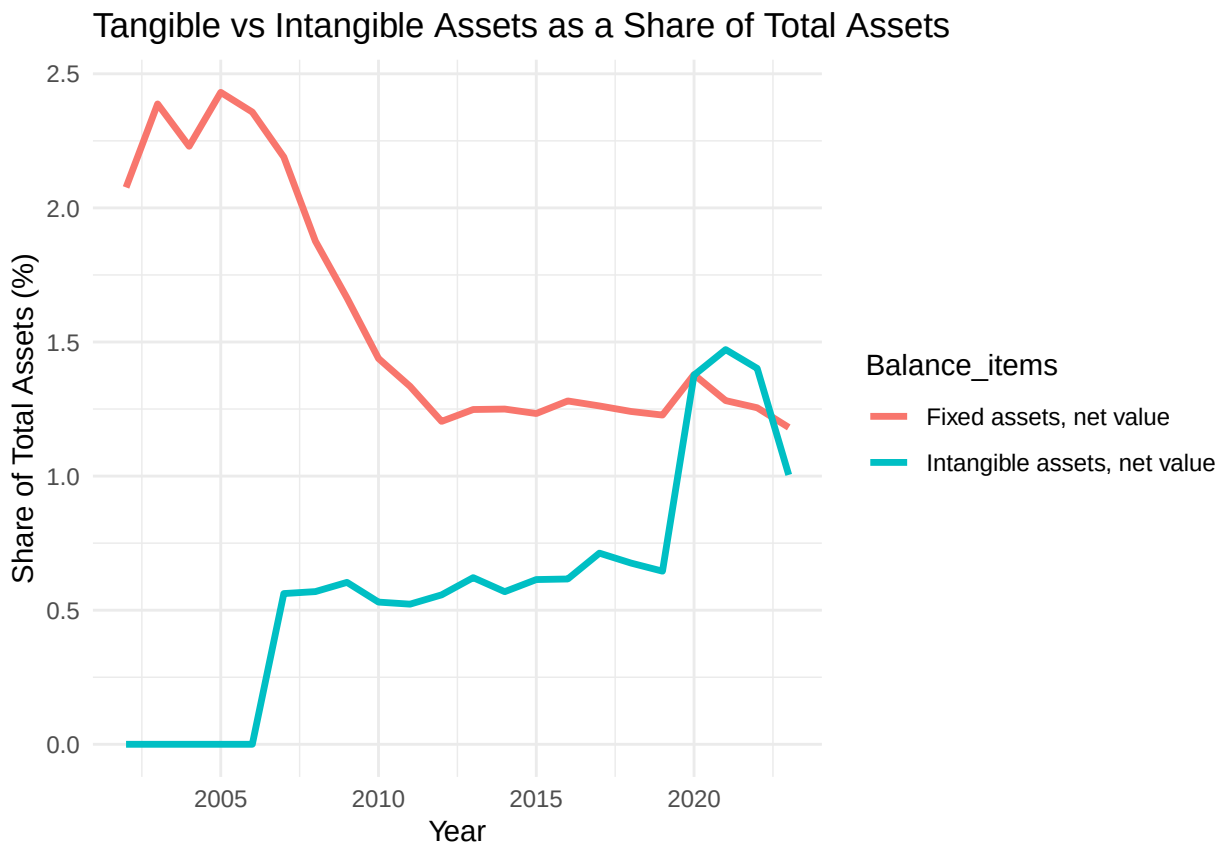
- Before a change in methodology in December 2012, most expenditures on intangible assets, such as spending on research and development (R&D) or organizational capital, were treated as intermediate expenses in the Canadian System of National Accounts rather than as capital investments.

  The only exceptions at that time were spending on software and mineral exploration, which were already treated as investments. This meant that the majority of intangible assets were considered inputs consumed during the production process, not as assets that contribute to economic output over an extended period.

- Meaning as per the authors devised cases. The Government has deemed Case 2 an accurate representation to the classification of intangible assets.

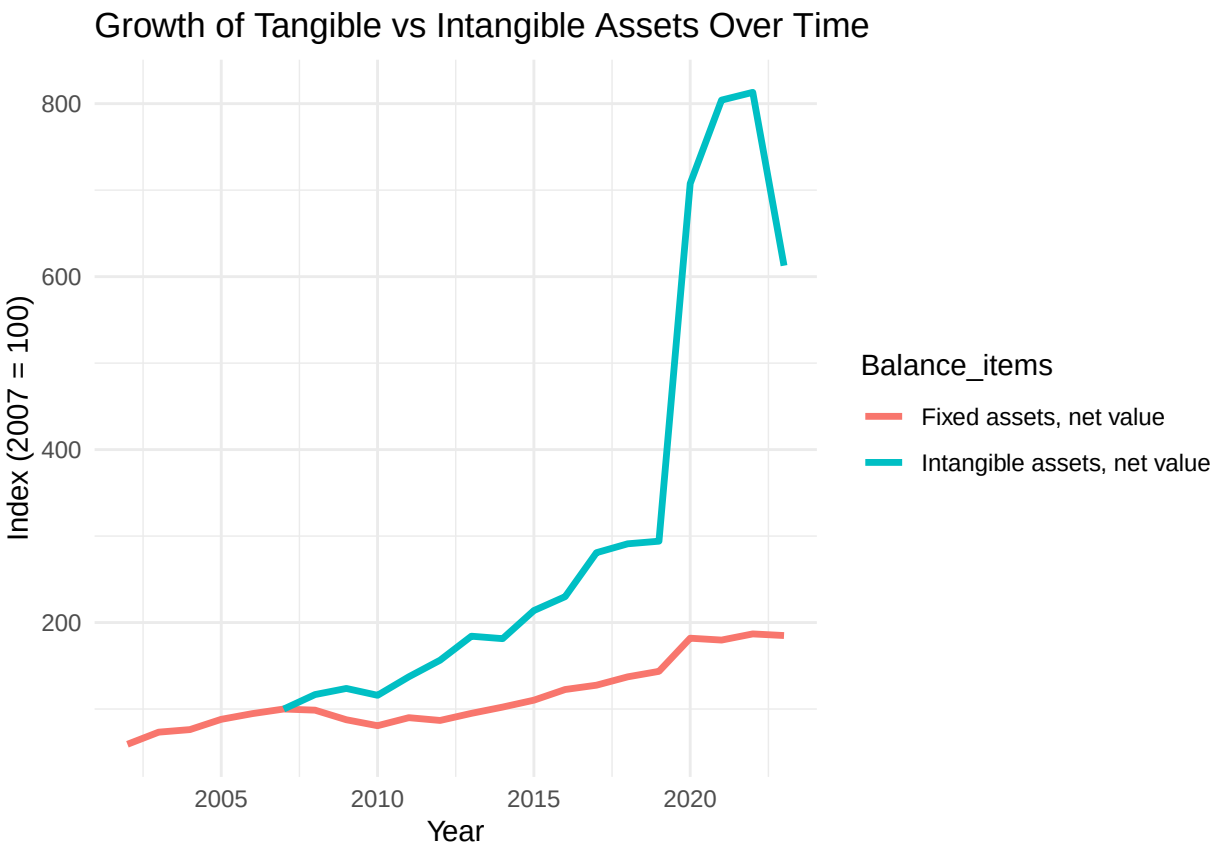
```
assets_df <- df %>%
  filter(
    Balance_items %in% c("Total assets", "Fixed assets, net value", "Intangible assets, net value"),
    GEO == "Canada"
  ) %>%
  group_by(Year, Balance_items) %>%
  summarise(VALUE = sum(VALUE, na.rm = TRUE), .groups = "drop") %>%
  group_by(Year) %>%
  mutate(TotalAssets = VALUE[Balance_items == "Total assets"]) %>%
  filter(Balance_items %in% c("Fixed assets, net value", "Intangible assets, net value")) %>%
  mutate(Share = VALUE / TotalAssets * 100) %>%
  ungroup()


ggplot(assets_df, aes(x = Year, y = Share, color = Balance_items)) +
  geom_line(size = 1.2) +
  labs(y = "Share of Total Assets (%)",
       title = "Tangible vs Intangible Assets as a Share of Total Assets") +
  theme_minimal()
```



These values in this figure are likely under represented.

```
df %>%
  filter(Balance_items %in% c("Fixed assets, net value", "Intangible assets, net value"),
         Industry == "Total, finance and insurance industries") %>%
  group_by(Balance_items) %>%
  mutate(Index = VALUE / VALUE[Year == as.Date("2007-01-01")] * 100) %>%
  ggplot(aes(x = Year, y = Index, color = Balance_items)) +
  geom_line(size = 1.2) +
  labs(y = "Index (2007 = 100)",
       title = "Growth of Tangible vs Intangible Assets Over Time") +
  theme_minimal()
```



These results are indexed to the year 2007. we can see how much asset types have grown over time.

Replicating Graphs From The Assigned Paper.

Looking back on the assignment, "Using the Statistics Canada report, prepare a one-page submission that includes". So I should probably include some graphics using the datatables from the paper. "Data, Intangible Capital and Economic Growth in Canada"

```
# This is a completely different section of the project. so lets clean up the environment and continue
rm(list = ls())

readRenviron("~/.Renviron")
data_path <- Sys.getenv("project_path")

file_path <- file.path(data_path, "Data, Intangible Capital and Economic Growth in Canada.xlsx")
sheet_names <- excel_sheets(file.path(data_path, "Data, Intangible Capital and Economic Growth in Canada

# This is the extracted data from the paper. I stuck it into an Excel workbook. Imported each sheet int
list_of_dfs <- lapply(sheet_names, function(sheet_name) {
      read_excel(file_path, sheet = sheet_name)
  })

#The data sources are named according to where they were found in the paper (i.e. The dataframe named "

Table_3 <- list_of_dfs[[1]]
```

```
Table_4 <- list_of_dfs[[2]]
Table_5 <- list_of_dfs[[3]]
Table_6 <- list_of_dfs[[4]]
# After combing through the data, and looking at the assignment evaluation criteria, i think i can hit
Table_7 <- list_of_dfs[[5]]

Chart_1 <- list_of_dfs[[6]]
Chart_3 <- list_of_dfs[[7]]
Chart_4 <- list_of_dfs[[8]]
Chart_5 <- list_of_dfs[[9]]
Chary_6 <- list_of_dfs[[10]]

rm(list_of_dfs)


# There are 4 entries in our "Industry" Column which display an aggregate of certian sectors and indust
#Total business sector
#Other goods-producing industries
#Manufacturing industries
#Service-producing industries

# We wish to remove these
Table_7 <- Table_7 %>% filter(Industry != "Total business sector")
Table_7 <- Table_7 %>% filter(Industry != "Other goods-producing industries")
Table_7 <- Table_7 %>% filter(Industry != "Manufacturing industries")
Table_7 <- Table_7 %>% filter(Industry != "Service-producing industries")


# The names are too long to be presentable. so let's work on a solution to shorten them.
head(Table_7$Industry)
```

```
[1] "Crop and animal production (11A)"
[2] "Forestry and logging (113)"
[3] "Fishing, hunting and trapping (114)"
[4] "Support activities for agriculture and forestry (115)"
[5] "Oil and gas extraction (211)"
[6] "Mining and quarrying (except oil and gas) (212)"
```

```
# I'm not liking the results for this function.
head(abbreviate(Table_7$Industry, minlength = 5, method = "both.sides", use.classes = F))
```

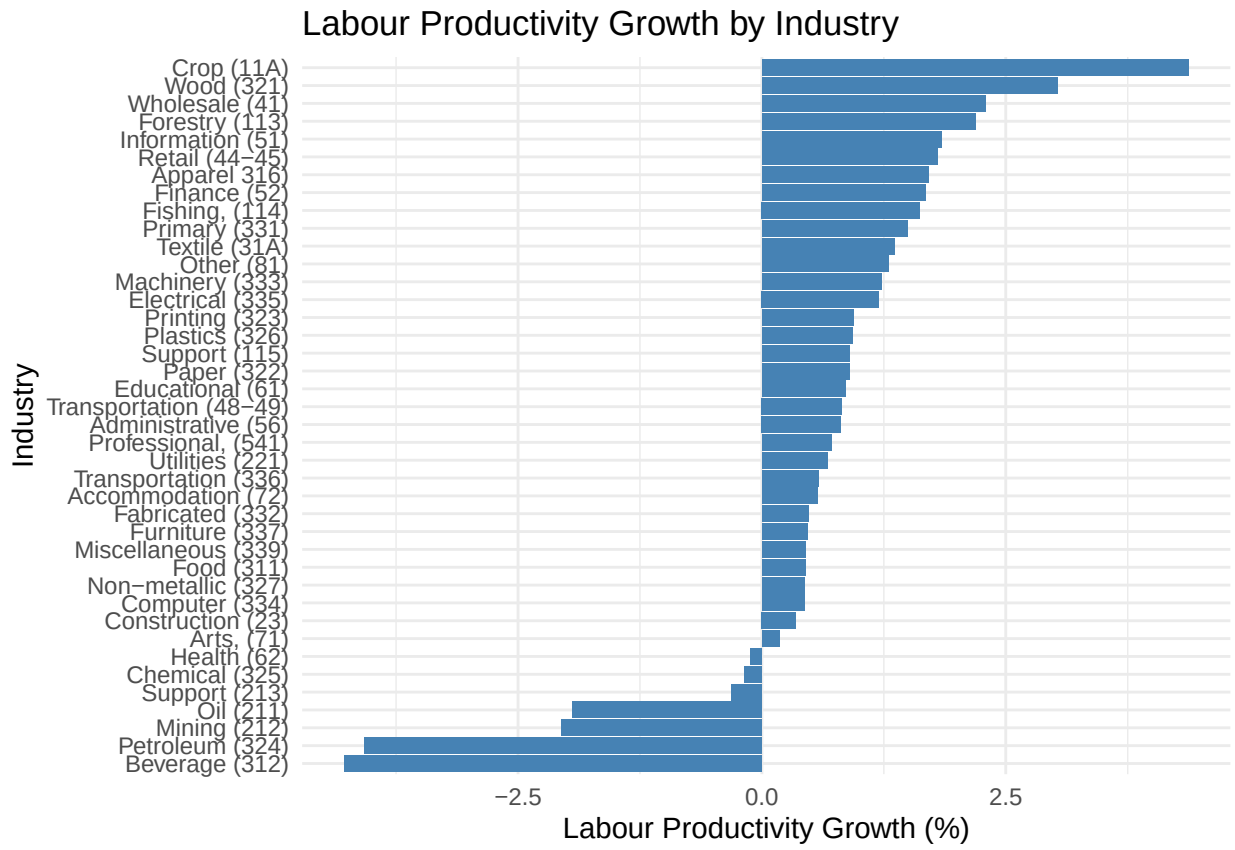
```
                     Crop and animal production (11A)
                                              "Caap("
                           Forestry and logging (113)
                                              "Foal("
                  Fishing, hunting and trapping (114)
                                              "Fhat("
Support activities for agriculture and forestry (115)
                                             "Safaaf("
                          Oil and gas extraction (211)
                                              "Oage("
      Mining and quarrying (except oil and gas) (212)
                                             "Maq(oag("
```

```r
library(stringr)

# keeps first word in these entries
Table_7$Industry_first <- word(Table_7$Industry, 1)

# This keeps only the first and last word. meaning "Crop and animal production (11A)" turns into "Crop
Table_7$Industry_first_last <- paste(
  word(Table_7$Industry, 1),
  word(Table_7$Industry, -1)
)
```

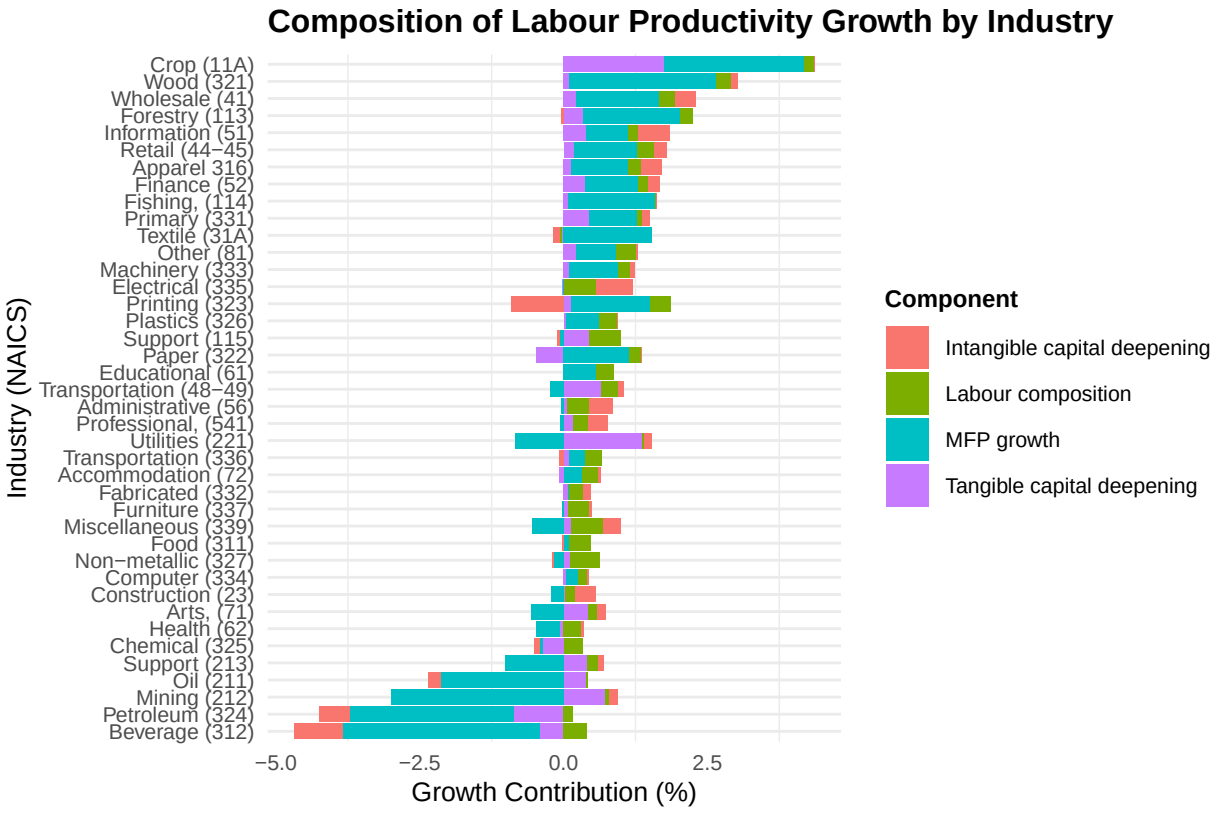
```r
# Bar chart
ggplot(Table_7, aes(x = reorder(Industry_first_last, `Labour productivity growth`),
                    y = `Labour productivity growth`)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(title = "Labour Productivity Growth by Industry",
       x = "Industry", y = "Labour Productivity Growth (%)") +
  theme_minimal()
```




```r
# Stacked bar chart
Table_7_long <- Table_7 %>%
  pivot_longer(cols = c(`Tangible capital deepening`,
                        `Intangible capital deepening`,
                        `Labour composition`,
```

```
                        `MFP growth`),
                names_to = "Component", values_to = "Value")

Stacked_Bar <- ggplot(Table_7_long, aes(x = reorder(Industry_first_last, `Labour productivity growth`),
                          y = Value, fill = Component)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Composition of Labour Productivity Growth by Industry",
       x = "Industry (NAICS)",
       y = "Growth Contribution (%)",
       caption = "Source: Statistics Canada. Data, Intangible Capital and Economic Growth in Canada. Ana
  theme_minimal() +
  theme(
    axis.text.y = element_text(size = 8), # y-axis labels (industry names)
    axis.text.x = element_text(size = 8), # x-axis labels
    axis.title = element_text(size = 10), # axis titles
    plot.title = element_text(size = 12, face = "bold"), # plot title
    legend.text = element_text(size = 8), # legend labels
    legend.title = element_text(size = 9, face = "bold"), # legend title
    panel.grid.major.x = element_blank() # remove default x grid
  )
Stacked_Bar
```



Economic Growth in Canada. Analytical Studies Branch Research Paper Series

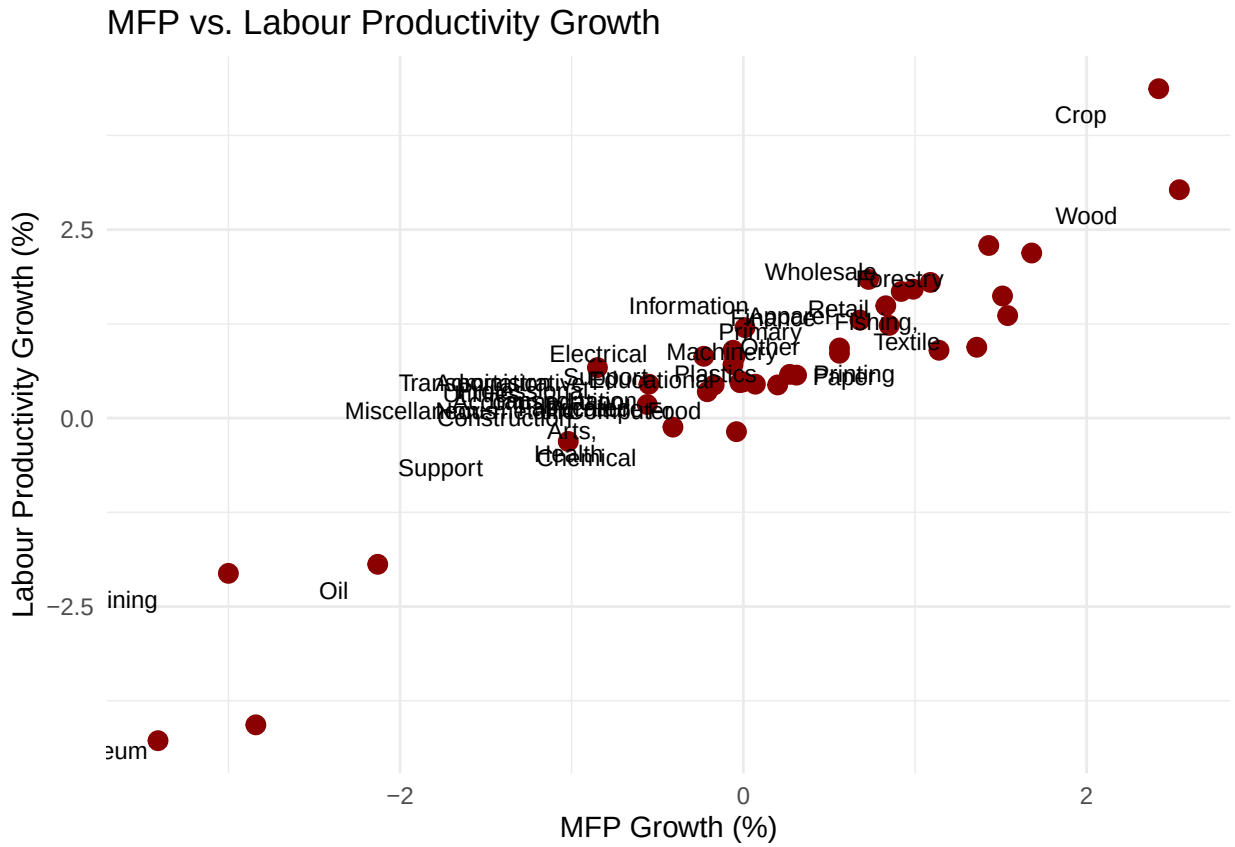
```
    ggsave("Stacked Bar Plot.jpeg", plot = Stacked_Bar, width = 12, height = 6, dpi = 300)
```

Understanding how labour productivity is affected by tangible and intangible assets, with considerations to MFP Growth, and Labour Composition.

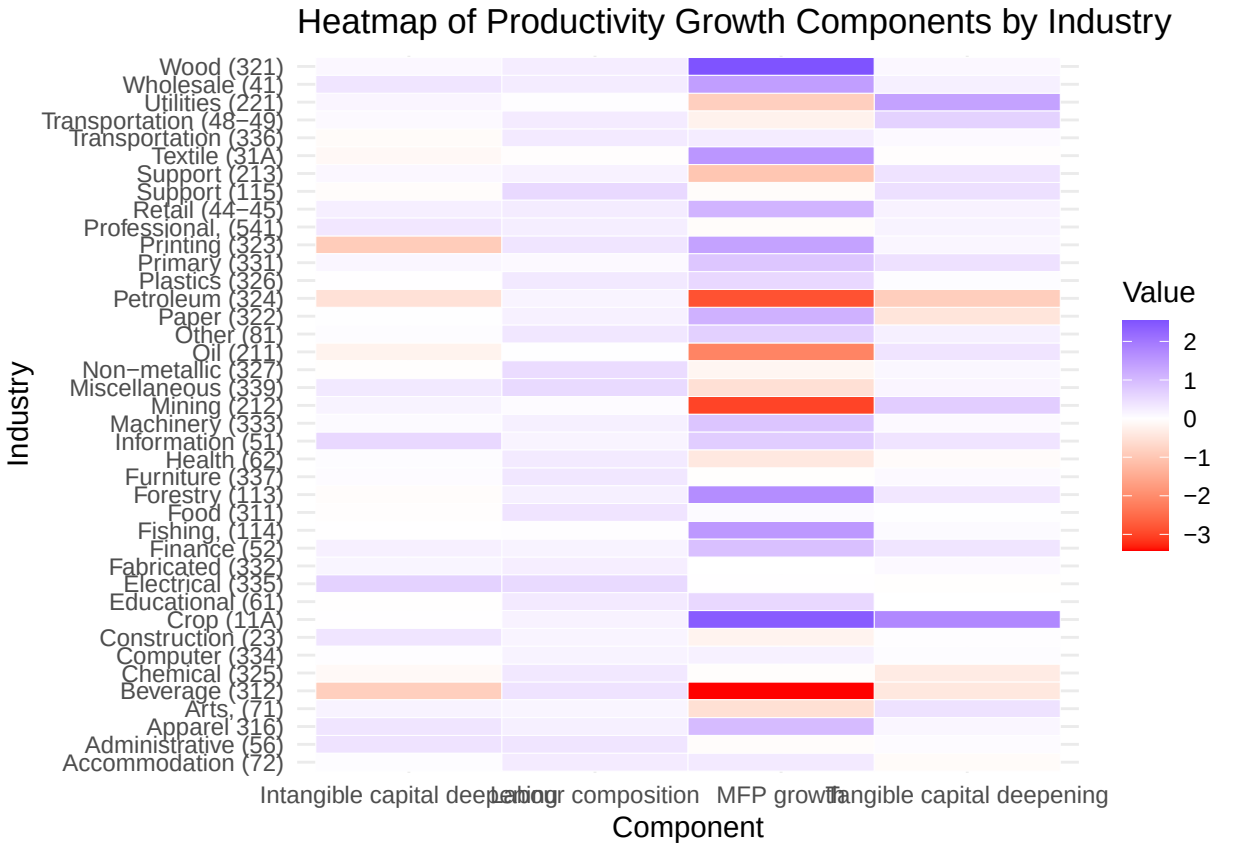
In essence from this graph, some industries benefit more from intangible assets.

Industries like Information, Retail, Apparel, Finance, and Machinery. Are more "Labour Productive" when high ammounts of intangible capital are acquired and invested into.

```
# Scatter plot
ggplot(Table_7, aes(x = `MFP growth`, y = `Labour productivity growth`, label = Industry_first)) +
  geom_point(color = "darkred", size = 3) +
  geom_text(hjust = 2, vjust = 2, size = 3) +
  labs(title = "MFP vs. Labour Productivity Growth",
       x = "MFP Growth (%)", y = "Labour Productivity Growth (%)") +
  theme_minimal()
```




```
# Heat map
ggplot(Table_7_long, aes(x = Component, y = Industry_first_last, fill = Value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "red", high = "blue", mid = "white", midpoint = 0) +
  labs(title = "Heatmap of Productivity Growth Components by Industry",
       x = "Component", y = "Industry") +
  theme_minimal()
```



Citations

Corrado, C., Haskel, J., Jona-Lasinio, C., & Iommi, M. (2022). Intangible Capital and Modern Economies. The Journal of Economic Perspectives, 36(3), 3–28. https://www.jstor.org/stable/27151252

Statistics Canada. Table 33-10-0499-01 Balance sheet, income statement and taxation statistics with selected financial ratios, by financial industries. DOI: https://doi.org/10.25318/3310049901-eng

Statistics Canada. (2024). Rural Canada Business Profile, balance sheet items of incorporated small businesses by industry and location indicator; profit margin-based calculations (Record ID: c1f40c2d-83fd-4d85-9dbc-d20ba326ac40) [Data Set]. https://open.canada.ca/data/en/dataset/c1f40c2d-83fd-4d85-9dbc-d20ba326ac40

Statistics Canada. (2024). Rural Canada Business Profile, balance sheet items of incorporated medium businesses by industry and location indicator; profit margin-based calculations (Record ID: 48d6f6c4-cb2f-42b8-a2a8-74da4ad9e1ab) [Data Set]. https://open.canada.ca/data/en/dataset/48d6f6c4-cb2f-42b8-a2a8-74da4ad9e1ab

Depreciated Results.

Any research project would be incomplete without getting caught up in a series of seemingly unrelated tasks, often necessary to complete an initial, simple task. These tasks here have been performed on the following datasets.

https://open.canada.ca/data/en/dataset/c1f40c2d-83fd-4d85-9dbc-d20ba326ac40

https://open.canada.ca/data/en/dataset/48d6f6c4-cb2f-42b8-a2a8-74da4ad9e1ab

Upon arriving at the point to create some graphs. I had realized there was no seperated value for intangible and tangible assets.

Data Sources

The Statistics Canada report "Data, Intangible Capital and Economic Growth in Canada" (2025) focuses on the measurement and growth contributions of intangible assets (e.g., data, R&D, software). It does not provide provincial or firm-level granularity. To contextualize these findings for Ontario's economy, we can draw on Rural Canada Business Profile datasets available from the Open Government Portal. These data sets include balance sheet items for incorporated medium businesses, reported by industry and location, with calculations based on revenues and profit margins.

There are two data options.

Revenue based analysis and profit based analysis. In the context of this assignment. Revenue-based plots tend to show the scale of industries/sectors in rural Canada. While profit-based plots provide efficiency/productivity differences. (could possibly link back to intangible capital adoption.) I have opted for the latter.

Importing Data

```
# using this hides the file paths. as this doc is posted on my GitHub [https://github.com/MatthewAndary
# if you wish to replicate my findings please feel free to refer to the citations for the data sources.
readRenviron("~/.Renviron")
data_path <- Sys.getenv("project_path")
small_business_raw <- read.csv(file.path(data_path, "33100587.csv"))
medium_business_raw <- read.csv(file.path(data_path, "33100599.csv"))


#Small Business data set
small <- small_business_raw

#Medium Business data set
medium <- medium_business_raw

# Removing our Raw data. won't be using it.
rm("medium_business_raw")
rm("small_business_raw")

glimpse(medium)
```

```
Rows: 85,540
Columns: 18
$ REF_DATE                                          <int> 2017, 2017, 2017~
$ GEO                                               <chr> "Canada", "Canad~
$ DGUID                                             <chr> "2021A000011124"~
$ North.American.Industry.Classification.System..NAICS. <chr> "All industries"~
$ Location.indicator                                <chr> "Rural and small~
$ Incorporation.status                              <chr> "Incorporated", ~
$ Business.characteristics                          <chr> "Total assets", ~
```

```
$ UOM                                                  <chr> "Dollars", "Doll~
$ UOM_ID                                               <int> 81, 81, 81, 81, ~
$ SCALAR_FACTOR                                        <chr> "thousands", "th~
$ SCALAR_ID                                            <int> 3, 3, 3, 3, 3, 3~
$ VECTOR                                               <chr> "v1430252969", "~
$ COORDINATE                                           <chr> "1.1.1.2.1", "1.~
$ VALUE                                                <dbl> 7255, 9388, 4465~
$ STATUS                                               <chr> "A", "", "", "",~
$ SYMBOL                                               <lgl> NA, NA, NA, NA, ~
$ TERMINATED                                           <chr> "", "", "", "", ~
$ DECIMALS                                             <int> 0, 0, 0, 0, 0, 0~
```

```
glimpse(small)
```

```
Rows: 1,295,448
Columns: 18
$ REF_DATE                                             <int> 2017, 2017, 2017~
$ GEO                                                  <chr> "Canada", "Canad~
$ DGUID                                                <chr> "2021A000011124"~
$ North.American.Industry.Classification.System..NAICS. <chr> "All industries"~
$ Location.indicator                                   <chr> "Rural and small~
$ Incorporation.status                                 <chr> "Incorporated", ~
$ Business.characteristics                             <chr> "Total assets", ~
$ UOM                                                  <chr> "Dollars", "Doll~
$ UOM_ID                                               <int> 81, 81, 81, 81, ~
$ SCALAR_FACTOR                                        <chr> "units", "units"~
$ SCALAR_ID                                            <int> 0, 0, 0, 0, 0, 0~
$ VECTOR                                               <chr> "v1429005314", "~
$ COORDINATE                                           <chr> "1.1.1.2.1", "1.~
$ VALUE                                                <int> 1065300, 1602200~
$ STATUS                                               <chr> "A", "", "", "",~
$ SYMBOL                                               <lgl> NA, NA, NA, NA, ~
$ TERMINATED                                           <chr> "", "", "", "", ~
$ DECIMALS                                             <int> 0, 0, 0, 0, 0, 0~
```

Recoding & Renaming Our Variables

```
#Recoding our year as a Date. helps with time series analysis
medium$Year <- as.Date(paste0(medium$REF_DATE, "-01-01"))
small$Year <- as.Date(paste0(small$REF_DATE, "-01-01"))


# Renaming for conciseness
medium <- medium %>% rename(Industry = North.American.Industry.Classification.System..NAICS.)
small <- small %>% rename(Industry = North.American.Industry.Classification.System..NAICS.)


# Currently the original Location.indicator varible has only two levels. "Rural and small town area" an
medium <- medium %>%
  mutate(Location = case_when(
    Location.indicator == "Rural and small town area" ~ "Rural",
```

```
    Location.indicator == "Functional urban area" ~ "Urban",
    TRUE ~ Location.indicator
  ))

medium$Location <- as.factor(medium$Location)

# Same operation with smaller business.
small <- small %>%
  mutate(Location = case_when(
    Location.indicator == "Rural and small town area" ~ "Rural",
    Location.indicator == "Functional urban area" ~ "Urban",
    TRUE ~ Location.indicator
  ))
small$Location <- as.factor(small$Location)
```

Removing Unnecessary Variables

```
# Keeping only these variables
small_clean <- small %>%
  dplyr::select(Year, Location,GEO, Industry, Business.characteristics, UOM, SCALAR_FACTOR, VALUE, DECI

medium_clean <- medium %>%
  dplyr::select(Year, Location,GEO, Industry, Business.characteristics, UOM, SCALAR_FACTOR, VALUE, DECI

# Removing these datasets from environment. in the spirit of keeping it clean.
rm(medium)
rm(small)
```

Value and Scalar Factor.

```
# Currently our values are scaled by a factor of 1000. meaning when Value = 1000. it's real value is =

# Scaling to original value
small_clean <- small_clean %>%
  mutate(Value_unscaled = VALUE*1000)

# Joining our decimals and unscaled values.
small_clean <- small_clean %>%
  unite("Dollars",Value_unscaled:DECIMALS, remove = FALSE, sep = ".")

# cleaning
small_clean <- small_clean %>%
  dplyr::select(-SCALAR_FACTOR, -UOM, -Value_unscaled)


# Same as above.
medium_clean <- medium_clean %>%
  mutate(Value_unscaled = VALUE*1000)
```

```r
medium_clean <- medium_clean %>%
  unite("Dollars",Value_unscaled:DECIMALS, remove = FALSE, sep = ".")

medium_clean <- medium_clean %>%
  dplyr::select(-SCALAR_FACTOR, -UOM, -Value_unscaled)

unique(medium_clean$Business.characteristics)
```

```
 [1] "Total assets"
 [2] "Total assets, first quartile"
 [3] "Total assets, second quartile"
 [4] "Total assets, third quartile"
 [5] "Total assets, fourth quartile"
 [6] "Total current assets"
 [7] "Total current assets, first quartile"
 [8] "Total current assets, second quartile"
 [9] "Total current assets, third quartile"
[10] "Total current assets, fourth quartile"
[11] "Accounts receivable"
[12] "Accounts receivable, first quartile"
[13] "Accounts receivable, second quartile"
[14] "Accounts receivable, third quartile"
[15] "Accounts receivable, fourth quartile"
[16] "Closing inventory"
[17] "Closing inventory, first quartile"
[18] "Closing inventory, second quartile"
[19] "Closing inventory, third quartile"
[20] "Closing inventory, fourth quartile"
[21] "Other current assets"
[22] "Other current assets, first quartile"
[23] "Other current assets, second quartile"
[24] "Other current assets, third quartile"
[25] "Other current assets, fourth quartile"
[26] "Net tangible and intangible assets"
[27] "Net tangible and intangible assets, first quartile"
[28] "Net tangible and intangible assets, second quartile"
[29] "Net tangible and intangible assets, third quartile"
[30] "Net tangible and intangible assets, fourth quartile"
[31] "Long term assets"
[32] "Long term assets, first quartile"
[33] "Long term assets, second quartile"
[34] "Long term assets, third quartile"
[35] "Long term assets, fourth quartile"
[36] "Total liabilities"
[37] "Total liabilities, first quartile"
[38] "Total liabilities, second quartile"
[39] "Total liabilities, third quartile"
[40] "Total liabilities, fourth quartile"
[41] "Total current liabilities"
[42] "Total current liabilities, first quartile"
[43] "Total current liabilities, second quartile"
[44] "Total current liabilities, third quartile"
[45] "Total current liabilities, fourth quartile"
```

```
[46] "Current bank loans"
[47] "Current bank loans, first quartile"
[48] "Current bank loans, second quartile"
[49] "Current bank loans, third quartile"
[50] "Current bank loans, fourth quartile"
[51] "Other current liabilities"
[52] "Other current liabilities, first quartile"
[53] "Other current liabilities, second quartile"
[54] "Other current liabilities, third quartile"
[55] "Other current liabilities, fourth quartile"
[56] "Long term liabilities"
[57] "Long term liabilities, first quartile"
[58] "Long term liabilities, second quartile"
[59] "Long term liabilities, third quartile"
[60] "Long term liabilities, fourth quartile"
[61] "Total equity"
[62] "Total equity, first quartile"
[63] "Total equity, second quartile"
[64] "Total equity, third quartile"
[65] "Total equity, fourth quartile"
```

We now have our cleaned datasets. We can move onto our graphs.

We have some problems. This dataset doesn't have the seperated values for tangible and intangible datasets. I've gone to find a new dataset.