

Assignment 1: Getting Started

Submission

Task 1: SDG Selection

1. The SDG which I selected is (10) **Reduced Inequalities**.
2. I chose the SDG **Reduced Inequalities** because tackling differences in wealth and opportunities is important for creating a fair and sustainable society. These gaps often lead to social tension and slow advancements in many fields such as healthcare, education, and economic development. Especially, after COVID-19, it triggers the largest increase in between-country inequality in three decades. I'm drawn to this goal since it confronts big global problems, and I think a visualization project can powerfully showcase these issues and possible ways forward.

Task 2: Visualization Analysis

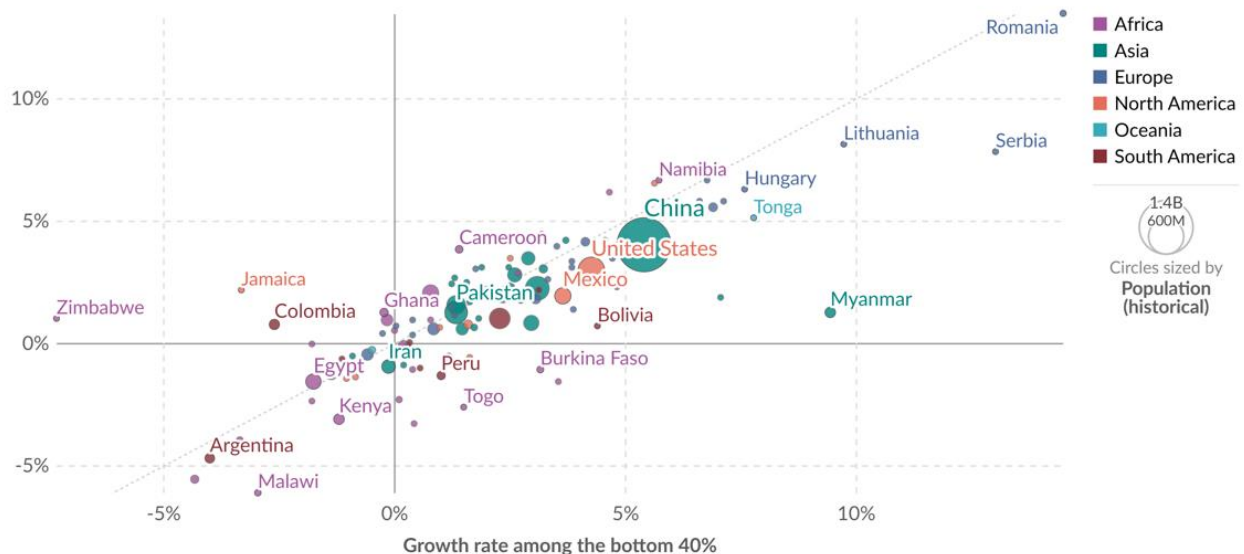
1. The link: <https://ourworldindata.org/sdgs/reduced-inequalities>

Annual growth of the income or consumption of the poorest 40% vs. the total population

Our World
in Data

The growth rate is calculated between two household surveys – the most recent survey available in 2022 and a survey falling approximately five years earlier. In countries below the dotted line, income or consumption growth is higher for the poorest 40% of the population than the national average.

Growth rate among the total population



Data source: World Bank

OurWorldinData.org/economic-inequality | CC BY

Note: Depending on the country and year, the data relates to income measured after taxes and benefits, or to consumption, per capita¹.

1. Per capita (income): "Per capita" here means that each person (including children) is attributed an equal share of the total income received by all members of their household.

This visualization directly relates to my **SDG Reduced Inequalities**, specifically Target 10.1, which aims to "by 2030, progressively achieve and sustain income growth of the bottom 40 per cent of the population at a rate higher than the national average." The scatter plot illustrates how countries are performing against this target by comparing the income or consumption growth of the poorest 40% to the overall population over time.

2. (a) Based on the lecture materials, Tufte's principles have graphical excellence, data-ink, chartjunk, and clear data visualization.

For graphical excellence, it gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space. This scatter successfully showing a direct comparison between two key metrics: growth for the poorest 40% versus the national average.

For maximizing the data-ink ratio, the proportion of data-ink versus the total ink used to print the graphics should be high. This visualization uses minimal non-data ink: the axes, a simple dotted line, and country dots are the focus.

For chartjunk, which is the interior decoration of graphics that does not tell the viewer anything new, this visualization avoids chartjunk effectively. The lack of excessive colors or unnecessary markers keeps the focus on the data points, avoiding the distractions.

Overall, this visualization aligns well with Tufte's principles.

(b) The generative-AI-based tool I use is **ChatGPT**. The prompt I use is "**Analyze my selected visualization as the SDG 10 Reduced Inequalities, starting from Tufte's principles.**"

ChatGPT's response:

To analyze the selected visualization—"Annual Growth of Incomes or Consumption of the Poorest 40% vs. the Total Population" from Our World in Data—through the lens of Edward Tufte's principles of data visualization, I'll apply his key concepts: graphical excellence, data-ink ratio, avoiding chartjunk, data density, and small multiples. These principles emphasize clarity, efficiency, and integrity in presenting data. Here's the analysis:

1. Graphical Excellence

Tufte defines graphical excellence as presenting complex ideas with clarity, precision, and efficiency while revealing the data's story.

- Strength: This scatter plot excels at conveying a complex relationship—comparing income growth rates of the poorest 40% versus the total population—across many countries in a single view. The use of a 45-degree reference line instantly communicates whether a country is reducing inequality (below the line) or not (above the line), aligning with SDG 10's focus on equitable growth. The interactive feature (available online) allows users to hover over data points for country-specific details, enhancing precision without cluttering the static view.

2. Data-Ink Ratio

Tufte advocates maximizing the "data-ink ratio," where the ink used in a graphic should primarily serve to represent the data, not decoration.

- Strength: The visualization uses minimal non-data ink. The axes are clearly labeled, the 45-degree line serves a functional purpose (as a benchmark), and the data points (dots) are the focal element. There's no unnecessary shading, 3D effects, or heavy gridlines that would detract from the data itself.

- Critique: The chart could slightly improve by reducing the thickness of the axes or simplifying the title font if it feels overly bold in the original design—though these are minor given the online format’s clean presentation.

3. Avoiding Chartjunk

Chartjunk refers to extraneous visual elements that distract from the data, such as excessive decoration or gratuitous effects.

- Strength: This visualization avoids chartjunk effectively. It lacks unnecessary embellishments like background patterns, over-the-top colors, or irrelevant icons. The color coding (if any, such as distinguishing regions) is subtle and purposeful, and the design remains utilitarian.

4. Data Density

Tufte encourages high data density—packing as much information as possible into a small space without overwhelming the viewer.

- Strength: The scatter plot achieves high data density by representing dozens of countries in one compact graphic. Each dot carries multiple pieces of information (country identity, growth rates for two groups), and the interactive version adds even more depth (e.g., years or specific values) without sacrificing readability.

- Critique: For a static version (like a printed image), some users might find densely packed dots hard to distinguish without labels, but the online interactivity mitigates this by allowing zooming or tooltips. This balance suits Tufte’s call for dense yet clear visuals.

5. Small Multiples (if applicable)

Tufte praises "small multiples"—repeating similar small visuals to show changes over time or across categories. While this specific chart doesn’t use small multiples in its static form (it’s a single scatter plot), the interactive version on Our World in Data allows users to filter by time periods or regions, effectively creating a dynamic equivalent.

- Strength: The ability to adjust the data by year or region online mimics small multiples by letting users compare subsets of the data (e.g., growth trends in Africa vs. Asia). This enhances understanding of how inequality reduction varies globally or temporally.

- Critique: A static version could benefit from explicit small multiples (e.g., side-by-side plots for different decades), but this would require more space and might dilute the immediate impact of the single-plot overview.

Overall Assessment:

From Tufte’s perspective, this visualization is highly effective. It achieves graphical excellence by clearly communicating a key metric of SDG 10—relative income growth—with precision and minimal clutter. The data-ink ratio is strong, with most elements serving the data rather than aesthetics. It avoids chartjunk, maintaining a clean, focused design, and offers high data density to reflect the global nature of inequality. While not a traditional small multiples display, its interactive features compensate by enabling deeper exploration.

Relation to SDG 10: Reduced Inequalities

The visualization embodies SDG 10's mission by spotlighting disparities in income growth, a core indicator of inequality. Its clarity and efficiency make it a compelling tool for raising awareness about where the world stands on reducing inequalities, aligning with Tufte's goal of letting data speak truthfully and powerfully.

Let me know if you'd like to dive deeper into any aspect or adjust this analysis!

(c) Overall, the response succeeded in delivering a clear, structured, and SDG-relevant analysis that adhered to Tufte's principles. However, it fell short in adaptability to different contexts, depth of critique, and descriptive richness.

(d) *My improved analysis:*

Tufte's principles of data visualization—graphical excellence, data-ink ratio, avoidance of chartjunk, and data density—emphasize clarity, efficiency, and integrity in presenting data.

For graphical excellence, it gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space. This scatter plot achieves that by plotting income growth of the poorest 40% against the national average for numerous countries. The 45-degree dotted line acts as a clear benchmark, instantly showing whether the poorest are gaining faster (below the line) or lagging (above it). This directly supports SDG 10's aim to reduce income disparities.

For data-ink ratio, Tufte urges maximizing ink used for data over decoration. Here, the visualization uses simple axes, a functional dotted line, and dots for countries, with little non-data ink.

For chartjunk, which is the interior decoration of graphics that does not tell the viewer anything new, this chart avoids such pitfalls, steering clear of excessive colors, 3D effects, or irrelevant markers. Its clean design keeps viewers focused on the data's story—inequality trends—without distraction.

For data density, Tufte values packing substantial information into small spaces. This visualization represents dozens of countries efficiently, each dot encoding multiple variables (country, two growth rates).

Overall, this visualization aligns closely with Tufte's principles, delivering a clear, data-rich view of income inequality trends with minimal clutter. It effectively highlights progress and gaps for SDG 10, making it a valuable asset for understanding global efforts to reduce inequalities.

Task 3: *Dataset Exploration*

1. ***The description of the dataset:*** The dataset tracks income inequality and redistribution across various countries from 1960 to 2023. It includes:
 - country: Country name.
 - year: Year of observation.
 - gini_disp: Gini coefficient of disposable income (post-tax and transfer), with standard error (gini_disp_se).
 - gini_mkt: Gini coefficient of market income (pre-tax and transfer), with standard error (gini_mkt_se).
 - abs_red: Absolute Gini reduction due to redistribution, with standard error (abs_red_se).
 - rel_red: Relative Gini reduction (%), with standard error (rel_red_se).

It covers numerous countries, with varying data availability, showing inequality before and after redistribution.

The link to the dataset: [swiid9_8.zip - Harvard Dataverse](#)

2. **The tool:** Python with Matplotlib and Pandas libraries.

My visualization:



Coding:

```
[2]: import matplotlib.pyplot as plt
      %matplotlib inline
      import numpy as np
      import pandas as pd
      import seaborn as sns
      import colorcet as cc
      %reload_ext autoreload
      %autoreload 2

[3]: gini_df = pd.read_csv('swiid9_8_summary.csv')

      # Filter Finland data and select relevant columns
      finland_df = gini_df[gini_df['country'] == 'Finland'][['year', 'gini_disp']]

      sns.set_style("white")
      # Plot the data
      plt.figure(figsize=(10, 5))
      sns.lineplot(data=finland_df, x='year', y='gini_disp', marker='o', color='r')

      # Customize the plot
      plt.title('Income Inequality in Finland (Gini Coefficient Over Time)', fontsize=14)
      plt.xlabel('Year', fontsize=12)
      plt.ylabel('Gini Coefficient (gini_disp)', fontsize=12)
      plt.show()
```

My visualization is a line chart showing the Gini coefficient for disposable income (gini_disp) in Finland over time, spanning from approximately 1970 to 2020. This suggests reduced inequality in the late 20th century (likely due to welfare policies) followed by an increase, possibly from economic changes or policy shifts. Finland maintains relatively low inequality compared to global standards.

My design choice: A red dotted line with circular markers highlights the trend and individual data points, ensuring clarity and precision. No gridlines or excess styling keep the focus on the data.

Maximizing the Data-Ink Ratio: My design uses minimal non-data ink (no gridlines, concise labels), focusing on the red line and markers to represent the Gini trend. This aligns with Tufte's principle, prioritizing data over decoration while retaining essential labels for clarity. The chart is simple and effective without cluttering.