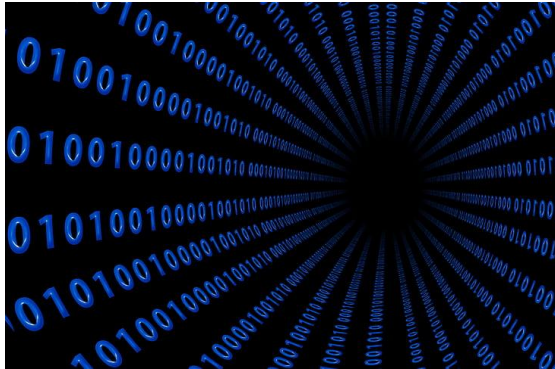


# Final Project: Personal Data Manifesto

Ruth Iiyambo

## Data

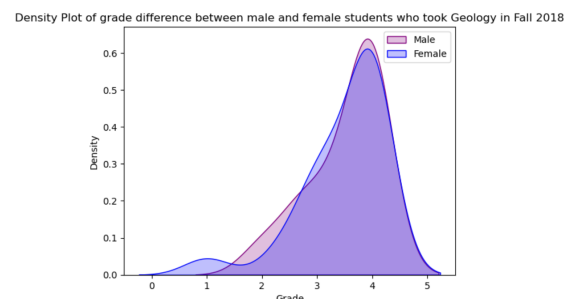
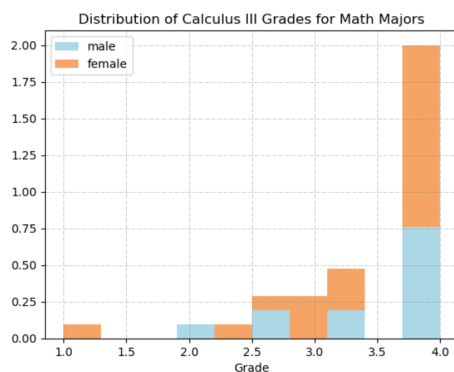


Data can be thought of as a collection of observations, measurements, or facts that are gathered and recorded for the purpose of analysis. While often perceived as raw, throughout the semester, we touched on how data is processed to some degree, however minimal, from the moment it is collected. Unlike information or knowledge, data does not inherently carry meaning; it exists as a potential from which meaning can be constructed.

In class, we have learned how to use Python and all its in-built functions to further process data to address different questions.

## What skills do you need to do data work:

Below are various data visualizations created during my coursework. These visuals demonstrate the technical skill to manipulate and present data effectively and also highlight patterns and insights. Each visualization serves as a means of data communication, and the importance of accessibility in the field of data science.



Showing the

Calculus grades for math majors male/female

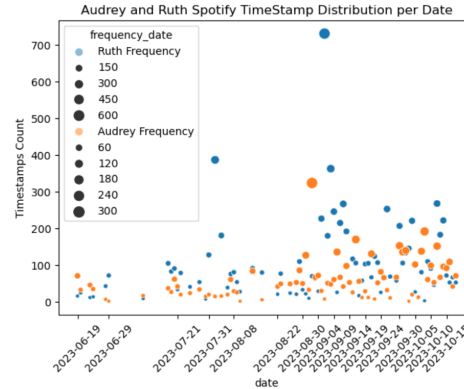
```

nam_resorts = soup.find_all('div', {'class': 'vp-portfolio_item-img'})
for item in nam_resorts:
    links = item.find_all('a') # Find all anchor tags
    for link in links:
        print(link.get('href')) # Print the href attribute of each anchor tag

dolomite-resort/
onkoshi-resort/
sossus-dune-lodge/
ai-ais-hotspring-and-spa/
gross-barmen-resort/
halali-resort/
hardap-resort/
namutoni-resort/
okaukuejo-resort/
papa-falls-resort/
waterberg-resort/
boplaas-campsite/
duwisib-castle/
hobas-lodge/
khorixas-restcamp/
miles-and-jakkalsputz/
naukluft-camp/
olifantsrus-campsite/
sesriem-campsite/
shark-island/
terrace-bay-resort/
torra-bay-campsite/

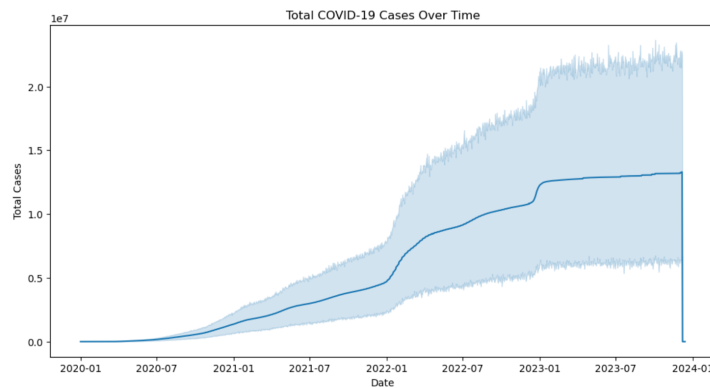
```

Web-scraping assignment and I worked on a website from Namibia showing different wildlife resorts.



A group project where we worked with personal data. This was our Spotify data.

## Problem-Solving through Data Analysis:



Data plays a crucial role in problem-solving. I will be using my cookbook as an example as I worked with COVID-19 data that at the time was a driving force for many of the decisions made. The power of data lies in its ability to reveal patterns and anomalies that might not be immediately visible. However, it's

essential to recognize both the strengths and limitations of data in this context.

One of the most significant capabilities of data is uncovering hidden issues like underreporting. For example, during a health crisis, consistent data analysis can reveal discrepancies in infection rates or mortality figures. If certain regions report significantly lower cases despite having similar conditions as high-incidence areas, this might suggest underreporting or testing inadequacies. Data can also highlight demographic disparities in disease impact, which might go unnoticed without thorough data analysis. The importance of data has been seen during many of the disease outbreaks such as cholera and COVID-19.

The reliability of data is dependent on the accuracy and comprehensiveness of the data collection process. Data might be skewed due to various factors, such as limited access to testing, reporting biases, and the inability to capture all cases. These limitations mean that while data can point toward certain trends or issues, it might not always provide a complete picture. This was seen in

my project 4 where I was working with “Access to electricity”. Some countries had data missing as there was no data available.

Beyond numerical analysis, data often falls short in capturing the full scope of human experiences. Quantitative data can tell us how many people are affected, but it doesn't capture the qualitative aspects such as the psychological impact on individuals, the social dynamics at play, or the lived experiences of those in different communities.

## Personal Influence in Data Science

Personally, data science stands at the forefront of knowledge creation in our increasingly “techy” world. My background, values, and perspectives play an important role in how I approach this field, especially in terms of ethical considerations and inclusivity.

Something that stood out for me during the semester is how each small bias in data projects all adds up and can lead to much bigger problems. I advocate for and strive to practice inclusive data science, where the views and experiences of these communities are actively integrated. This approach makes projects more robust and applicable to a broader spectrum of society.

I believe that context is as crucial as the data itself. Data is shaped by many factors. For instance, in analyzing economic data, I consider not just the numbers but also the socioeconomic conditions that might influence those numbers. This approach helps in creating more accurate, nuanced analyses that reflect real-world complexities. The risk of biases and stereotypes being perpetuated through data science is a significant concern. In my work, I'm vigilant about the potential for racial, gender, or socio-economic biases in the datasets I use. For example, in a health data project, I carefully examined the data to ensure it didn't inadvertently reflect or reinforce existing health disparities.

## Advice to someone who is hoping to become a data scientist

### Data Methodology

This document covers our methodology for the following datasets:

- Electricity **generation** (TWh)
- Electricity **net imports** (TWh)
- Electricity **demand** (TWh)
- Installed power generation **capacity** (GW)
- **Emissions** from electricity generation (Mt CO<sub>2</sub>e)

All data is available to download for free on the Ember website. It is provided on an 'as is' basis, and is assembled using the best data available to us at any given time. Every effort has been made to ensure accuracy, and where possible we compare multiple sources to confirm their agreement. We take no responsibility for errors.

If you notice an issue or have any suggestions, comments, or questions, please do contact us at [data@ember-climate.org](mailto:data@ember-climate.org).

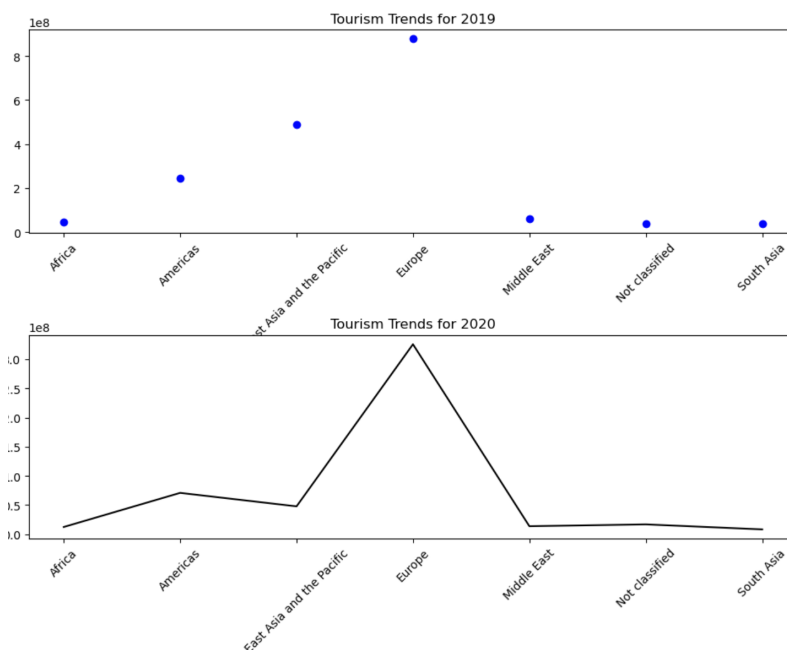
Honesty when working on these projects. Give credit when it is due, and provide evidence of why some data might not be an accurate representation based on assumptions that had to be made, to complete the project. For Project 5, the data we used was well explained.

Each term had an explanation of what it meant based on the study. For example, the study explicitly defined “Electricity as “safe cooking facilities and a minimum consumption: 250 kWh/year for rural households and 500 kWh/year for urban ones. By 2020, most continents had near-total electricity access, but only 5 of 54 African countries reached 100% access.” This is useful as people working with this data will know that it might not be that the country does not have excess to electricity, but rather that they do not meet the requirements of the study.

## Data Presentation:

Another important insight gained this semester centers on the various nature of data representation. The decision lies on us as data communicators to select the most effective method of visualization that integrates our intended message. To illustrate this point, I will include two contrasting visualizations in my report:

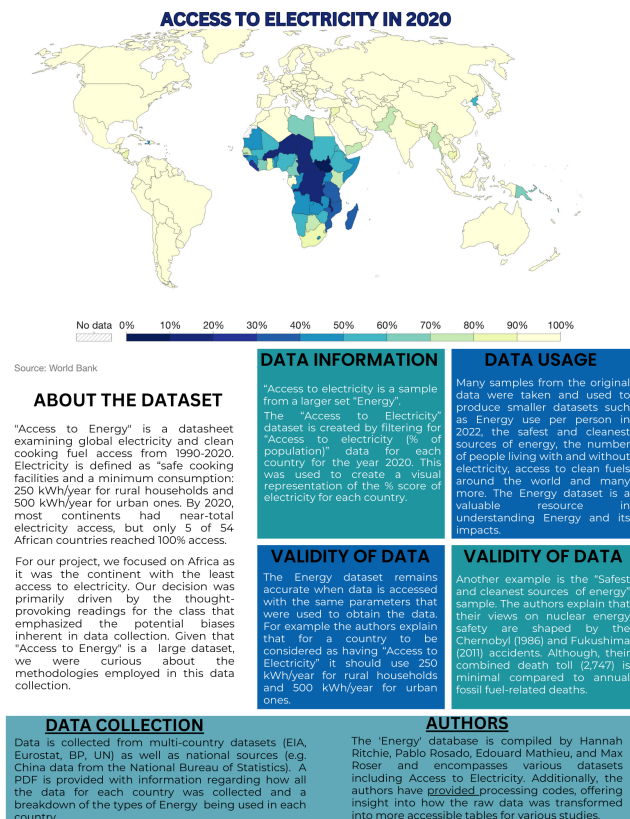
1. The first, is not as clear but still provides valuable insights into the data, this might be a little confusing for some people
2. The second is a model of more clarity, offering a more comprehensive interpretation of the data at a glance.



The part of data science extends beyond analysis; it includes the critical skill of storytelling through data. A study loses its value if its presentation fails to resonate with the audience. It shows that understanding by the analyst is not enough; comprehension by the audience is key. Tailoring visualizations to the audience's level of expertise or interest is not just beneficial, it's necessary for effective communication. This approach

ensures that our findings are not just understood by a select few but are accessible and informative to all, thereby bridging the gap between data and decision-making.

What kinds of problems can you solve / questions can you answer through data analysis and visualization? What kinds of questions can you not solve?



This flyer is a project we did for the class where I used the essential skill of distilling large data into concise, digestible segments. Being able to filter through the complexities of "raw" data to filter important information is crucial in the data analysis process. It involves not only the elimination of unnecessary details but also the organization and presentation of data in a way that highlights potential opportunities and insights.

I have learned that effective communication through data is about more than just simplification. It's about converting large data of numbers into a narrative that speaks clearly about trends, challenges, and opportunities. My project explored electricity access across different regions, requiring me to filter relevant statistics from global datasets, interpret their significance, and then visually translate them to address key questions about global electricity distribution and its

socio-economic implications. I became more thoughtful at making decisions about which types of visualizations would most effectively convey the story within the data.