# DATS 6401 Visualization of Complex Data – Final Project Assignment

## Project Goal

The goal of the project is to engage in a full-cycle analysis and development of data visualizations in order to answer at least three questions you have about some topic of your choosing. For this project, you can either work individually or in groups (**no more than three people**). You will design your visualizations, evaluate results, ask additional questions, and develop supporting visualizations.

Required Submissions:

* A written report
* Your data visualization product

## Topics

The topic you choose is open-ended, please choose a topic that you are interested in and are genuinely curious about. Think of some questions that you do not know the answer to.

## Data

The data can be pulled from multiple sources; it does not need to be a single dataset. Be sure to get data from the original source. For example, if you wish to work with data collected and distributed by the Centers for Disease Control, that is where you should go to access the data, not a third party that has posted the data. Avoid overused datasets (like the Titanic dataset) as well as those used in Kaggle (or similar) competitions.

Some examples are:

* [Open Data DC](https://opendata.dc.gov/)
* [Centers for Disease Control and Prevention (CDC)](https://www.cdc.gov/DataStatistics/)
* [NYC Open Data](https://opendata.cityofnewyork.us/)
* [US Bureau of Labor Statistics](https://www.bls.gov/" \t "_blank)

Datasets and data sources will need to be provided to me by 10/30 (Week 8) for approval. The dataset used must contain at the very least 1000 records, preferably in the several thousand range.

## Platform

Your project can be completed in Python, R, or Tableau. Projects developed primarily using code should be built in a notebook format with visuals displayed. Final submissions should be in a zip format with the data files in properly referenced file pathways so that the code will run smoothly.

## Submission Format

In a zip folder, please submit the following:

* Your written report in a Word Doc or PDF format, that includes the elements required in the Written Report category of the rubric (filter column A = Written Report).
* Your Data Visualization product, for example the .twb Tableau file and/or a .ipynb file. Make sure it includes all the elements required in the Data Visualization Product category of the rubric (filter Column A = Data Visualization Product)
* ALL source data files for the data visualizations -- if I cannot recreate your visualizations on my side, the project will not be accepted.

## Written Report Requirements

Your written report must follow the following outline:

### Introduction

Explain why you chose this topic, and the questions you are interested in studying. List at least three questions you intend to explore in your project.

List team members and a description of how each contributed to the project.

### Description of the Data

Describe how the data was collected, how you accessed it, and any other noteworthy features.

### Analysis of Data Quality

Provide a detailed, well-organized description of data quality, including textual description, graphs, and code. This should include variable names, a description of each variable, the data type of each variable, and descriptive statistics of each variable.

### Main Analysis

Provide a detailed, well-organized description of your findings, including textual description, graphs, and code. Your focus should be on both the results and the process. Include, as reasonable and relevant, approaches that didn’t work, challenges, the data cleaning process, etc. The length should be at least 500 words.

### Key Findings

Provide a short nontechnical summary of the most revealing findings of your analysis written for a nontechnical audience. The length should be approximately 250 words.

### Conclusion

Discuss limitations and future directions, lessons learned.

## Data Visualization Product Requirements

Your final data visualization product can take the following formats:

* Jupyter Notebook
* GitHub.io
* R Markdown
* Shiny Dashboard
* Tableau
* Other platforms with approval – please schedule office hours to review your alternative platform by 11/13 (Week 10), otherwise I will not accept alternative platforms.

### Requirements

* Provide at least 10 distinct charts.
* Use at least 5 distinct data visualization types (i.e., scatter plots, bar chart, heat map, data table etc.).
* Incorporate appropriate chart titles, axis labels, data descriptions where necessary, descriptions of charts, inferences that you make, and additional questions you can think of to explore further.

### Additional Information

The written report portion should to some extent also be incorporated into your data visualization product. The audience should be able to follow the data visualization product like an article or storyboard, being guided through each section with descriptions and observations that lead to the next section. Sections can be centered around a distinct visualization or multiple visualizations.

## Rubric

The project is made up of 100 points total and graded according to the following sections. A breakdown of the rubric is posted separately on Blackboard.

* Topic selection (originality)
* Topic selection (creativity)
* Evenly distributed contributions of each team member
* Written report
* Visualization (variety)
* Visualization (aesthetically pleasing)
* Visualization (usefulness)
* Visualization (clear and easy to use)
* Program and tools (interface)
* Program and tools (easy to use)
* Technical knowledge
* Clarity of concepts and analysis
* Focused? (easy to follow)
* Error free
* Presentation (organization)
* Presentation (professionalism)
* Presentation (quality)
* Key findings
* Conclusions

Note that a project that does not work or is fairly incomplete will not receive partial credit and will be given a grade of 0. The rubric applies only to working and complete projects that use complex datasets.

**Slide 1: Title Slide**

* **Title:** “World Population Trends: Insights and Predictions”
* **Subtitle:** Key Insights from Historical and Future Projections
* **Visuals:**
  + Relevant image or graphic (e.g., globe, population pyramid).
* **Details:**
  + Your name and team details (if applicable).
  + Date and institution or organization (if required).

**Slide 2: Introduction**

* **Title:** Introduction and Objectives
* **Content:**
  + Briefly explain the topic: Global population growth is a critical issue influencing economics, urbanization, and resource planning.
  + List objectives:
    1. Examine historical trends in global population distribution.
    2. Identify regions with the fastest and slowest growth.
    3. Predict future population dynamics and their implications.
  + Mention your interest in exploring population sustainability and challenges.
* **Visuals:**
  + A simple world map or demographic trend chart.

**Slide 3: Research Questions**

* **Title:** Key Research Questions
* **Content:**
  + How has the population growth of individual countries varied over time?
  + What factors contribute to the rapid population growth in some regions?
  + How are aging populations affecting developed countries?
  + What predictions can be made about future global population trends?
* **Visuals:**
  + Use a checklist or a question mark graphic to make it engaging.

**Slide 4: Description of Data**

* **Title:** Data Overview
* **Content:**
  + Briefly describe data sources, collection methods, and timeline (e.g., UN or World Bank datasets, 1950–2023).
  + Mention key features: global population data by year and country.
  + Data quality: consistent, reliable, covers all countries, but limitations exist in projections.
* **Visuals:**
  + Table summarizing data structure: Variables (Year, Country, Population, etc.).

**Slide 5: Population Growth Trends (Visualization)**

* **Title:** World Population Growth Over Time
* **Content:**
  + Highlight trends observed:
    - Rapid growth in Africa and Asia.
    - Slower growth in developed regions.
    - China's plateau and India’s growth.
  + Mention insights drawn from the graph.
* **Visuals:**
  + The **interactive world map visualization** (the uploaded graph). Include key observations in bullet points next to it.

**Slide 6: Regional Focus: Africa**

* **Title:** Africa’s Role in Global Growth
* **Content:**
  + Mention rapid increases in Nigeria, Congo, and Ethiopia.
  + Discuss higher fertility rates and improved healthcare contributing to this trend.
  + Predict Africa's share of global population by 2100 (~40%).
* **Visuals:**
  + A heatmap or bar chart showing Africa's growth compared to other regions.

**Slide 7: Declining Populations**

* **Title:** Population Decline in Developed Countries
* **Content:**
  + Examples: Japan, Russia, Eastern Europe.
  + Key reasons: aging population, low fertility rates, emigration.
  + Challenges: shrinking workforce, increased dependency ratio.
* **Visuals:**
  + Line graph or comparative chart of population decline trends.

**Slide 8: Key Findings**

* **Title:** Insights and Key Takeaways
* **Content:**
  + Population growth disparities: high growth in Africa, stagnation in developed countries.
  + Urbanization pressures: mega-cities in Africa and Asia to face infrastructure challenges.
  + Aging populations: policy responses in Japan and Europe.
* **Visuals:**
  + Infographic summarizing key findings.

**Slide 9: Predictions and Implications**

* **Title:** Predictions and Future Trends
* **Content:**
  + Population to peak at ~10.4 billion by 2080.
  + Africa's population to dominate growth.
  + Climate and resource challenges in high-growth areas.
  + Strategies for managing aging populations.
* **Visuals:**
  + Include a forecast chart or pyramid showing global population predictions.

**Slide 10: Challenges and Limitations**

* **Title:** Challenges in Population Analysis
* **Content:**
  + Incomplete data for certain regions.
  + Uncertainty in projections due to political or environmental factors.
  + Difficulty predicting migration patterns.
* **Visuals:**
  + A simple graphic of a balancing scale or a caution symbol.

**Slide 11: Conclusion**

* **Title:** Conclusion
* **Content:**
  + Summarize the importance of understanding population trends for global planning.
  + Restate key findings and predictions.
  + Emphasize future opportunities and risks.
* **Visuals:**
  + Include a quote or a powerful statistic (e.g., “By 2100, 1 in 3 people will live in Africa”).

**Slide 12: Acknowledgments and References**

* **Title:** Acknowledgments and References
* **Content:**
  + Thank collaborators, data sources, and institutions.
  + List references for datasets, graphs, and research papers.
* **Visuals:**
  + A simple thank-you graphic or institution logos.