A3. Exploratory Visual Analytics

<div style="border:1px solid #ccc; display:inline-block; padding:10px 20px;">Submit Assignment</div>

Due Nov 28 by 11:59pm   **Points** 20   **Submitting** a file upload   **File Types** pdf

---

Individual Assignment

Overview:

This assignment is about using visualizations to explore data and pose questions. Effectively translating data into visualizations to pose new questions and answer them is extremely valuable for human-centered designers.

In this assignment, you will use Tableau to perform *exploratory visual analysis* on a real-world dataset we will provide you. The **goal** is to gain practice formulating and answering questions through the creation of graphs during the process of visual analysis.

The really short version:

> Look at a dataset we provide → come up with questions → iteratively answer questions with visualizations → write up your process, describing your questions and corresponding visualizations

Detail and guidance (the long version):

You will work individually for this assignment. The write-up should not be more than 2500 words in length (~8-10 pages, including images). This is ~750 words per question/hypothesis, as well as a brief intro, data profile, and concise but thoughtful conclusion.

You will formulate two (2) compelling and appropriate **questions/hypotheses** that can be answered through a visual analysis of the **World Development Indicators** **(http://data.worldbank.org/data-catalog/world-development-indicators)** dataset. This data set is large, rich, and multivariate. Your goal is not only to pose thoughtful and creative questions (see below for important notes about formulating these) but to produce for each one, a finished visualization that exhibits excellence and integrity. *Please limit yourselves to one map-based visualization (or none at all, if you so choose)*.

You must keep a **record** of your process of formulating and addressing each question/hypothesis. You must capture the intermediary views/iterations that help you progress from your initial question, through to your finished visualization and description. To this end, **take plenty of notes and screenshots** to use in your final write-up!

The exploratory visual analysis is an **iterative** process, and you may need to backtrack, reformulate or revise your question, or try a new way of manipulating or presenting the data. During your exploration of the data, you should create and record various types of views, as appropriate for the data and question you are exploring, including (but not limited to): bar charts, scatter plots, maps, time-series, etc. Note how different views support different questions and may reveal areas for further questions or exploration.

What to Do:

1. **Download** the **World Development Indicators** **(http://data.worldbank.org/data-catalog/world-development-indicators)** data from **www.worldbank.org** **(http://www.worldbank.org)** (you can also create data extracts of your own by **visiting this link** **(http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators)**).
2. **Explore** the data set, and create a profile of the data to include in your writeup (see below for what a profile of the data should include).
3. **Formulate and capture** your initial two (2) questions or hypotheses about the data. We recommend formulating ones of various types such as: a comparative question, a correlation question, a geographically-oriented question, a time-related or trend question, etc.
4. **Create** various visualizations that attempt to answer or address your questions/hypotheses about the data. Record what does or does not work, why or why not, and how you progress through each one.
5. **Capture** screenshots of the graphs/charts you create along the way. Use them to support your descriptions of your process.
6. **Iterate** as needed. Refine questions. Try new views. Alter your approach. Repeat as necessary, until you feel you have arrived at a compelling answer in the form of a visualization - or reframed your question!
7. **Write up (use section headers!)** your process (including a final, appropriately labeled vis) and results for each of the three questions/hypotheses you have formulated. **Make sure you take any and every opportunity to use terms and concepts from the course to justify your reasons and claims.** This is the time to geek-out on jargon.
8. **Summarize** with a short (1 paragraph) **reflection** on your overall findings, and a concise description and critique of the tools you used at the end of your report.

For each question/hypothesis you will need to present the following in your writeup, **in this order**:

1. **The question/hypothesis you are addressing (clearly stated, final version);**
2. **A description and discussion** of your process;
3. **A final and well-designed visualization** that is appropriately labeled and captioned in accordance with the best practices we have learned in class, and;
4. **A short but thorough description** of how the visualization addresses the question, using **concepts and terms from class.**

Structuring your write-up

You must use the following sections/headers in your final write-up:

- Introduction
- Data Profile
- Q1
- Q2
- Discussion/Conclusion

Additional Information:

To profile a data set, briefly describe:

- **Where** did the data set come from (provenance)? What's in it?
- **How big** is data set (how many rows? how many variables? file size?).
- **What types** of data variables present? What are the dimensions/type?
- **What** is the overall perceived quality of the data? What's missing? What do you wish it included? Any noticeable outliers? Any other anomalous or curious things that jump out at you?

Notes on formulating good questions/hypotheses:

You are welcome to frame your inquiries as either answering a question or as trying to prove/disprove a hypothesis. For those that need a quick refresher (in most basic terms)...

- Example of a question: *what is the average airspeed velocity of an unladen swallow?*
- Example of a hypothesis: *I predict there is a correlation between an African swallow's airspeed velocity and their larger wingspan and lighter skeletal structure compared to European swallows.*

Be sure that the question/hypothesis *is not so simplistic* that it is not interesting or insightful, but not so open-ended that you are not able to effectively or reasonably address it. For example, if I were looking at 100 humans and multivariate data about them (height, weight, age, eye color, etc.):

- *Which person is the tallest?* - this is far too simplistic; it does not warrant visual exploration or have a good chance of revealing anything interesting about the data.
- *Which of these people is going to die next?* - While this is certainly an interesting (if not somewhat morbid) question, we have very little chance of being able to satisfactorily answer it in any meaningful or compelling way and is thus perhaps overly ambitious or open-ended.

Suggestions for possible tools you might use to examine the data and construct your visualizations (feel free to use whichever tools make sense to you):

For manipulating data: OpenRefine, Trifacta Wrangler, Excel, Python

For creating visualizations: Tableau, (you may also employ D3, Excel, Google Charts, Chart.js, Raw, Plot.ly, Carto, if justified)

See the assignment rubric for breakdown of points.

Please post questions about this assignment on the Slack channel.