CSC/SDS 235: Visual Analytics

Fall 2024

HW 01: Exploratory Data Analysis

This is a group assignment (3-4 students) – I recommend choosing collaborators with complementary skillsets to yours!
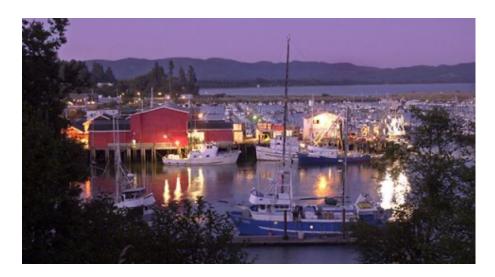
Goals:

- Explore realistic data via visualization
- Specify next-step action suggestions based on data insights

Instructions

Overview

Scenario

Welcome to Alderwood, a fictitious town in central Washington state! The year is 2003. Alderwood was once a popular vacationland American town. A beautiful community situated on the banks of the pristine Alderwood River, it is famous for excellent fishing, hiking and other recreational activities. Alderwood's fine water quality and lovely countryside are renown throughout the region.

Unfortunately, following record levels of growth, the "dotcom" industry has recently collapsed. The teetering US economy has entered a recession, heavily impacting the technology sector, which in turn raised unemployment levels in California, Oregon, and Washington to uncomfortably high levels. Tourism to many parts of the region dropped off severely.

In addition to tourism, the rural area surrounding Alderwood has long been supported by the beef and dairy cattle industry. In addition to the tourism decline, routine testing at a local farm discovered bovine spongiform encephalitis (BSE, also known as "mad cow disease"). This incident severely damaged beef sales, both nationally and internationally, resulting in a further depressed economy.

But good news! A famous scientist has recently relocated to Alderwood to lead the unexpected opening of a new private laboratory facility specializing in agri-business. The residents hope this could be the locus of a business renaissance in the area, and the influx of young talent to the area has caused quite a stir.

But something seems fishy... A not-so-reliable source has stated that high-paying, high-tech employment is now "a sure thing" in Alderwood, claiming unwavering support from "the high-rolling big wigs at City Hall". Another informant overheard that "if you support the bosses, the bosses will look after you." Backroom political dealings may be undermining efforts to boost the local economy, and it's up to you and your team to get to the bottom of it!

The Data

To support your analysis, you have been granted access to:
- 1000+ news stories from the Alderwood Daily News, plus a few other items collected by the previous investigators
- a spreadsheet with voter registry information
- a phone call log
- maps of Alderwood and the surrounding area

The data is all available via the course website. You may use as much or as little of the data in your analysis as you like. Some is useful, other parts might be less so... but you won't know what's what until you explore!

Your Objective

Flex your growing visual analytic muscles to take over the investigation of Alderwood's ongoing shenanigans. In this exploratory data analysis challenge, you may use any tools you have at your disposal to generate hypotheses, identify evidence, and formulate next-step action suggestions for additional information-gathering to present to your 'supervisor' (Ab).

Want some help getting started? Here are a few **questions** to guide your analysis:

- What is the **ongoing situation** in this scenario?
- Who are the **relevant players**? Which are innocent bystanders, and which are deliberately engaged in deceptive activities? How are the relevant players connected?
- What is the **time frame** in which this situation unfolded?
- What **locations** were relevant to the plot?
- What **events** occurred during this time frame? (*Note: Events are things that occur in a short, discrete time frame, such as "moving to campus".*)
- What **activities** were going on in this time frame? (*Note: Activities occur over a longer span of time, such as "going to college".*)

Don't worry too much about getting the "right answer" - instead, focus on making sure that the evidence you present supports your hypotheses of what the roles and relationships are and the motivations of the person(s) involved.

Good luck!

Acknowledgements

The dataset used was prepared by Jereme Haack, Carrie Varley, Mark Whiting and Katie Wolf from Pacific Northwest National Laboratory as part of the VAST Challenge 2006.

Deliverables

You will submit *four* deliverables for this assignment:

1. Sketches of the visualization(s) you intended to create.
2. The write up you would present to your supervisor based on your analysis.
3. Code (and a README.txt with instructions for running the code) that generates the visualization(s) in your write up.

4. A reflection (the entire group can write a reflection together, or group members may write individual reflections) that includes:
   - How each group member contributed to the final submission
   - One obstacle you encountered and how you overcame it
   - If you were to do this assignment again, what you would do differently.

Submission

Submit your deliverable(s) in a .zip on Gradescope. If you worked on the reflection as a group, submit as a group (https://guides.gradescope.com/hc/en-us/articles/21863861823373-Adding-Group-Members-to-a-Submission), otherwise submit (all pieces) individually.

Rubric

The following matches the rubric you will see on Gradescope. **Note your sketches and reflection weight most heavily into your grade.**

	Missing / Not Complete (0)	Approaching (3)	Meets (5)	Exceeds (6)
Sketches	Not submitted or not readable.	Sketches are difficult to read and/or need more detail. They do not demonstrate appropriate visual mappings (as discussed in lecture), or clearly support the analysis objectives.	Sketches are difficult to read and/or need more detail. They include some appropriate visual mappings (as discussed in lecture), but not all visual mappings are appropriate. Some visualizations do not support the analysis objectives.	Sketches are detailed, clear, and easy to read. They demonstrate appropriate visual mappings (as discussed in lecture), and clearly support the analysis objectives.
Reflection	Not submitted or not readable.	Reflection does not fully address all three points listed above. And/or needs improvement in one or more of the following areas: formatting, grammar and spelling, clear, concise writing.	Reflection addresses all three points listed above, but answers are not thoughtful. It is well formatted, contains good grammar and spelling, and clear, concise writing.	Reflection thoughtfully addresses all three points listed above. It is well formatted, contains good grammar and spelling, and clear, concise writing.

	Missing / Not Complete (0)	Approaching (1)	Meets (2)
Code	Not submitted.	Code does not run.	Code runs.
Write-up	Not submitted or not readable.	Write up addresses some but not all the objective(s) of the assignment. It could use improvement in one or more of the following areas: formatting, grammar and spelling, clear, concise writing. Hypotheses are unclear and/or not supported by visualizations shown.	Write up clearly addresses the objective(s) of the assignment. It is well formatted, contains good grammar and spelling, and clear, concise writing. Hypotheses are present and well supported by visualizations shown.