

Assignment 1

URBAN 520: Quantitative Methods in Urban Design and Planning

Due: October 20, 2026 11:59 pm Pacific Time

Total: 70 points, plus 5 points of extra credit

- You need to submit the following via Canvas: a PDF or HTML report (handwritten answers may be scanned into a PDF) with the answers in order, together with all files needed to reproduce your results, such as .Rmd, SPSS syntax (.sps), an Excel file (.xlsx), or a Tableau packaged workbook (.twbx).
- Each student is expected to submit their own assignment. Study groups are allowed, but you are expected to acknowledge the names of your collaborators, along with a short description of the types of collaboration, at the beginning of each submission.
- The only AI tool you may use for this assignment is UW's Purple AI. You should treat AI as a human collaborator instead of directly asking for answers or copying solutions. If you use AI in completing the assignment, you need to acknowledge how you used it, cite it appropriately, and submit the exported TXT conversation history.
- Whatever you submit, you must be able to explain how you did it and why. We may ask you to walk us through any part of your work at any time, and being unable to explain your own submission will be treated as if the work were not yours.
- You will have 6 penalty-free late days in total for assignments (max 3 late days per submission). Any submission delayed beyond the first 3 days will be penalized 10% per day for that specific assignment, but those days will not count toward your used late days.

Question 1 Short Answer

(20 points)

These questions should be answered briefly (one or two sentences) without using calculators.

- A. (2 points) The following table shows on-time arrival rates data from the U.S. Department of Transportation (DOT). How many observations are in the data? How many variables? What are the levels of measurement for those variables?

Table 1: On-Time Arrival Rates by Carrier, July 2026

Carrier	Airports Reported	On-Time Arrivals (%)	Rank
Envoy Air	154	79.8	1
Alaska Airlines	85	75.9	2
SkyWest Airlines	241	74.6	3
Delta Air Lines	143	71.9	4
United Airlines	131	71.5	5
Allegiant Air	120	68.6	6
Republic Airways	76	68.1	7
Southwest Airlines	106	68.0	8
American Airlines	124	65.0	9
Frontier Airlines	79	62.5	10
PSA Airlines	97	59.1	11
JetBlue Airways	71	55.7	12

Source: U.S. DOT, Air Travel Consumer Report, Table 1A, July 2026.

- B. (2 points) Both of the following lists have the same average of 50. Which one has a larger standard deviation and why? No calculation needed.
- 25, 30, 40, 50, 60, 70, 75
 - 25, 30, 40, 50, 60, 70, 75, 90, 10
- C. (2 points) A study shows that the average miles driven per person per day in a city is 8 miles, with an SD of 5 miles. What are the average and SD in kilometers? (1 mile $\approx$ 1.6 kilometers)
- D. (2 points) Which number would be larger: the median home sale price in Seattle or the average home sale price in Seattle, and why?
- E. (2 points) (True or False) Vector and raster data can be converted into each other without information loss if we use a proper GIS analysis tool.
- F. (2 points) A tract's median rent rose from \$1,000 to \$1,210 over two years. What is the compound annual growth rate (CAGR), and why is $21\% \div 2$ the wrong answer?
- G. (2 points) The UW campus tract has a Socioeconomic Factors Rank of 9 out of 10 (higher = more impacted) on the Washington State EHD Map based on Washington State Environmental Health Disparities Map. Is the tract disadvantaged in the sense the indicator is meant to capture?

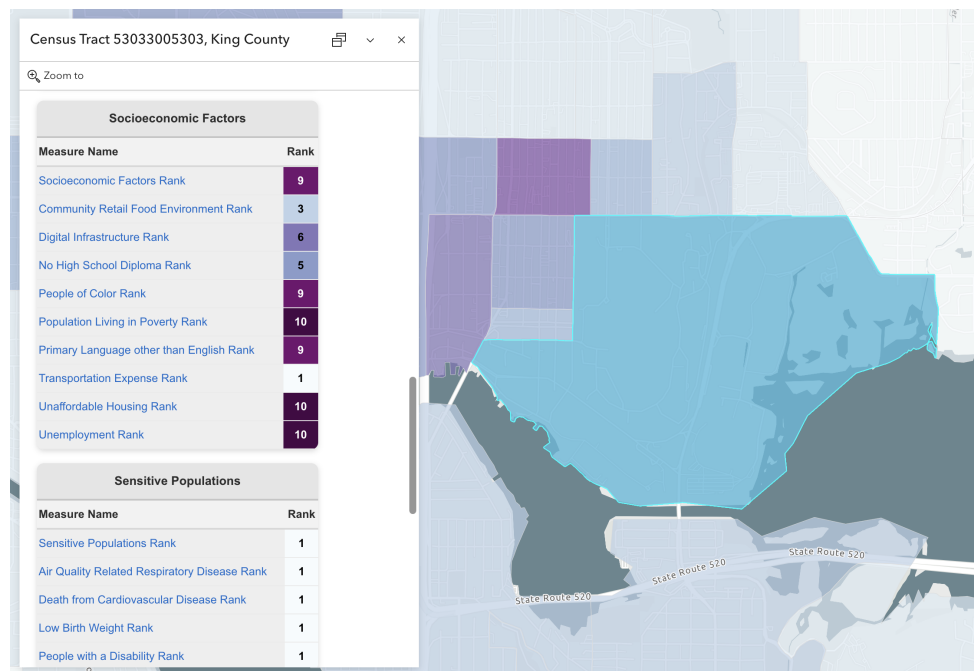


Figure 1: UW Campus in WA Environmental Health Disparities Map

- H. (2 points) The following table shows the Census Bureau's estimates of median household income for five states. Specify the level of measurement of each of the three variables. A classmate averages the five medians and reports the result as the median household income of these five states. Is that right? Would weighting each state by its number of households fix it?
- I. (2 points) A planner wants to know median household income changes in each Seattle census tract between 2022 and 2024. They take the ACS 2022 five-year estimate as the 2022 value, the ACS 2024 five-year estimate as the 2024 value, and subtract. Is it correct and what would you do?
- J. (2 points) The city moves one tract from District A to District B. No one's income changes, but average household income rises in both districts. How does it happen? Give a simple example.

Table 2: Median Household Income for Sample States, 2023

GEOID/FIPS Code	State	Estimate (\$)
08	Colorado	95,641
21	Kentucky	62,914
31	Nebraska	76,781
41	Oregon	82,515
50	Vermont	83,597

Source: 2023 American Community Survey (in 2024 Inflation-Adjusted Dollars), U.S. Census Bureau.

Question 2 Critique of Data Visualizations

(5 points)

As we learn more about data visualization, we have begun to discuss some general principles. For this section, please find at least one example of a visualization that impresses you, either in a good or bad way. Please attach the example image, give the link/source, and:

1. Describe what the visualization is trying to communicate;
2. Discuss what it does and does not do well using the principles we talked about in class;
3. Share on Canvas (https://canvas.uw.edu/courses/1923201/discussion_topics/10757428);
4. (Optional) Comment on two visualizations posted by classmates.

A few probing questions to think about:

1. Is the data visualization static, animated, or interactive? Narrative or exploratory?
2. What kind of purpose do you think it serves? How would you order its objectives (appeal, comprehension, and retention)?
3. Does it contain the basic elements of a graph (e.g., title, axis labels, encoding)? Is anything missing or not clear to the audience?
4. What kind of encoding does it use (e.g., position, length, slope, angle, area, color hue, et.)? How precisely can viewers read these encodings?
5. What kind of organizational principles do you observe (e.g., hierarchy, unity, balance, grouping, and spacing)? Are any of these principles missing?
6. Does the visualization meet your needs as a viewer (e.g., autonomy, relatedness, and competence)?

Question 3 Environmental Health Disparities in Washington

(30 points)

Washington State has a tool similar to California's CalEnviroScreen, named the Environmental Health Disparities (EHD) Map, maintained by the Washington State Department of Health (DOH). Version 3.0 was just released in 2026, and DOH is accepting public comments until the end of October 2026.

Table 3: Data Sources and Documentation

Item	Note
EHD Map	Browse tracts and measures.
Full EHD Version 3.0 Extract	Full dataset.
EHD Map Report, Version 3.0	Technical report (April 2026).
EHD Map Version 3.0 Updates	Changes from version 2.0 (February 2026).
Washington State Institute for Public Policy Technical Review	Review of version 2.0 (2022).

A. (5 points) The Structure of the Data

1. How many rows does the full EHD Version 3.0 Extract dataset have?
2. Which census geography (2010 or 2020 tracts) does Version 3.0 use?
3. Which variable(s) can be used as unique ID(s) / identifier(s)?
4. Check the EHD Version 3.0 technical documentation: what is the level of measurement for the raw value of *Ozone Concentration*?
5. Check the Washington State spatial data portal; can you find how many tracts the tool identifies as *Highly Impacted Census Tract* without downloading the data?

B. (10 points) **Making and using the Indicators**

1. Describe the path from the concept to the final indicator in no more than 200 words.
2. The report states that the model behind the EHD map is based on CalEnviroScreen. Compare EHD Version 3.0 with CalEnviroScreen 5.0. Name at least 2 differences and explain how it could change which tracts are designated as overburdened or disadvantaged.
3. Many Washington laws and policies rely on this map. Check RCW 70A.02.080, RCW 70A.65.230, and try to find more program referred this map. Briefly discuss: how the EHD data enters the law, policy, or program, and what is at stake (such as funding allocation).
4. (Optional) Anything else you find interesting or important.

C. (15 points) **Exploratory Visualization**

Choose at least two variables (raw values, not the ranks), each from a different theme. These variables should be interesting to you, and you should expect them to be related. Then, conduct some exploratory visualization. The purpose is to explore variables and their relationship, which means you don't have to make every figure perfect.

1. For each variable:
 - (a) Produce a histogram (state the bin width and your reasons) and one other figure (e.g., boxplot, cumulative distribution);
 - (b) Report the mean, median, and standard deviation;
 - (c) Briefly discuss the distribution of each variable (e.g., shape, spread, outliers, missing data).
2. For two or more variables:
 - (a) Produce a scatterplot with a simple trend line;
 - (b) Produce at least one other display suitable to your data (e.g., bubble chart, heatmap);
 - (c) Briefly describe any relationship you find.

Question 4 Creative and Deceptive Visualization

(15 points)

- A. (9 points) In this section, you will select one central theme or message based on your previous analysis (or from another dataset of your choice). Then, explore how different visual design choices can communicate it. You should try to produce one **perfect** figure:
1. Think about actions/targets of visualization.
 2. Choose different visualization idioms rather than default plots whenever possible.
 3. Experiment with more formatting and customization options to creatively express your message.
 4. Demonstrate conceptual clarity (a clear story or insight).
 5. Write a brief explanation of the message to the viewer and key design choices.
- B. (6 points) Building on your previous figure, intentionally create a deceptive version of the same visualization: one that misleads the audience while still appearing legitimate at first glance. Think about the plausible authors, audiences, and the decision the deceptive figure is meant to push.

Your challenge is to subtly distort perception without using obvious tricks such as removing data or adding false numbers. For instance, you might manipulate scales, reorder categories, or change

visual emphasis in ways that guide the viewer toward an incorrect interpretation. In your write-up, briefly reflect on what design decisions make the visualization deceptive, and how small graphical changes can influence interpretation.

Question 5 Extra Credit

(5 extra credit points)

- A. (1 points, extra credit) How many hours did you spend on this assignment?
- B. (2 points, extra credit) Show that $\sum_{i=1}^n (x_i - \bar{x}) = 0$.
- C. (2 points, extra credit) Show that for every c ,

$$\sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2.$$

Conclude that $\bar{x}$ minimizes the sum of squared deviations.