Fall 2024- ISTE 782: Visual Analytics　　　　　　**Rochester Institute of Technology**

Assignment 1

Question 1.(30 points) There are 6 graph below. For each graph below answer questions:

a. What type of chart is it? What does it visualize?

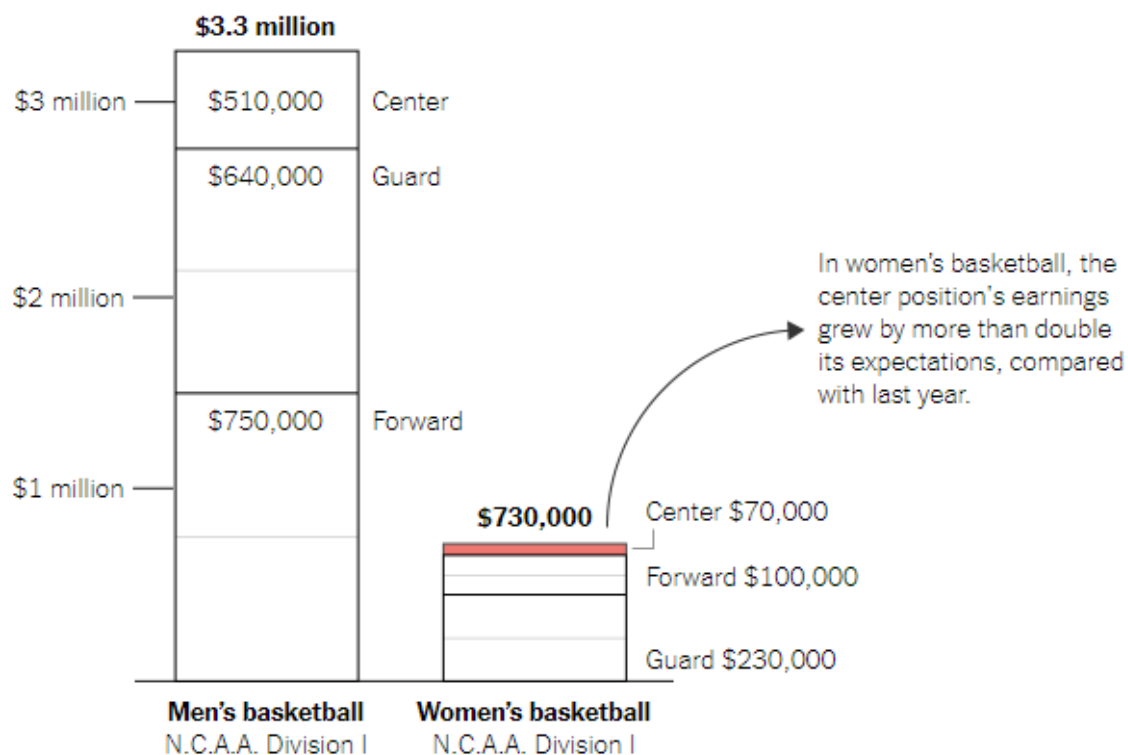(For example, it is a pie chart; it shows part of a whole.)

b. Based on Wilke's definitions and the lecture slides, evaluate the charts below in terms of what aspects are good, bad, ugly, or incorrect regarding data visualization. Why do you think these aspects stand out? Provide an explanation in your own words.

c. If the graph is accurate and visually appealing, explain what it represents (Hint: consider the x and y axes, legends, titles, and labels). If the graph appears to have issues, describe how you would fix it. You can either sketch the corrected version by hand or provide a detailed explanation of how it can be improved. **(5 points)**
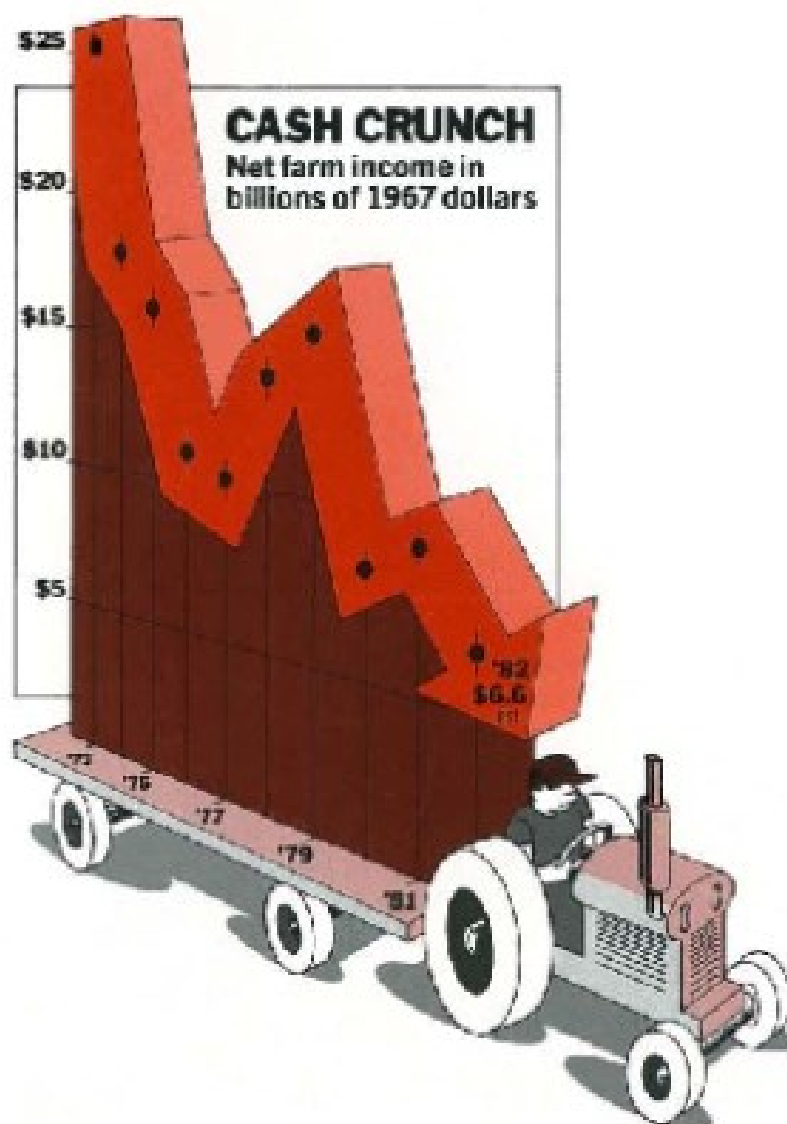
1) (5 points)

How much top-earning basketball players make

Expected annual compensation for players, on average, by position

$3.3 million

$3 million —	$510,000	Center
	$640,000	Guard
$2 million —		
	$750,000	Forward
$1 million —		

In women's basketball, the center position's earnings grew by more than double its expectations, compared with last year.

$730,000 — Center $70,000

Forward $100,000

Guard $230,000

Men's basketball
N.C.A.A. Division I

Women's basketball
N.C.A.A. Division I

Source: Opendorse. Data is based on N.I.L. transactions disclosed through or processed by Opendorse between July 1, 2021, and June 30, 2024. · Note: To be included in the calculations players' earnings must rank in the top 25 at their position.
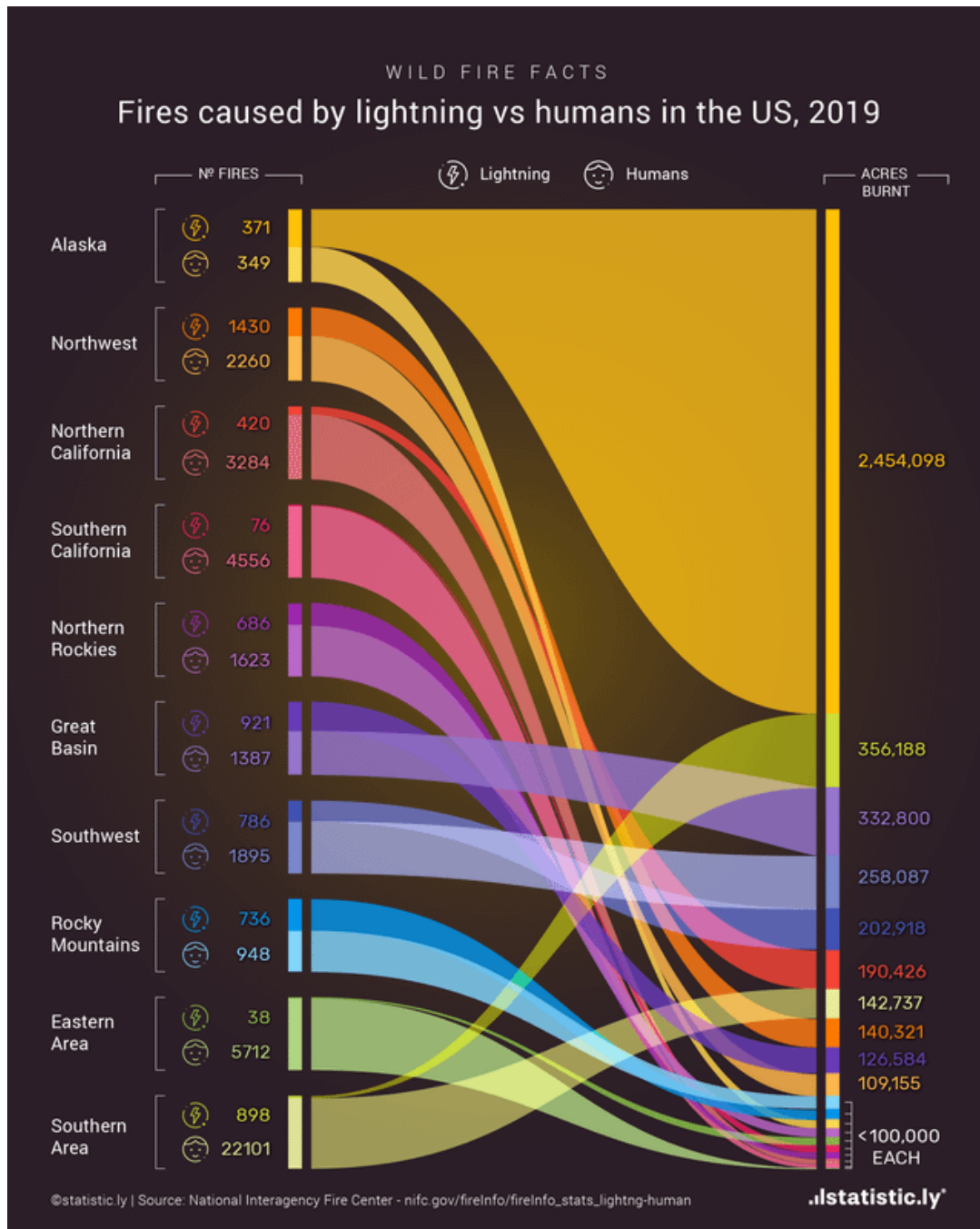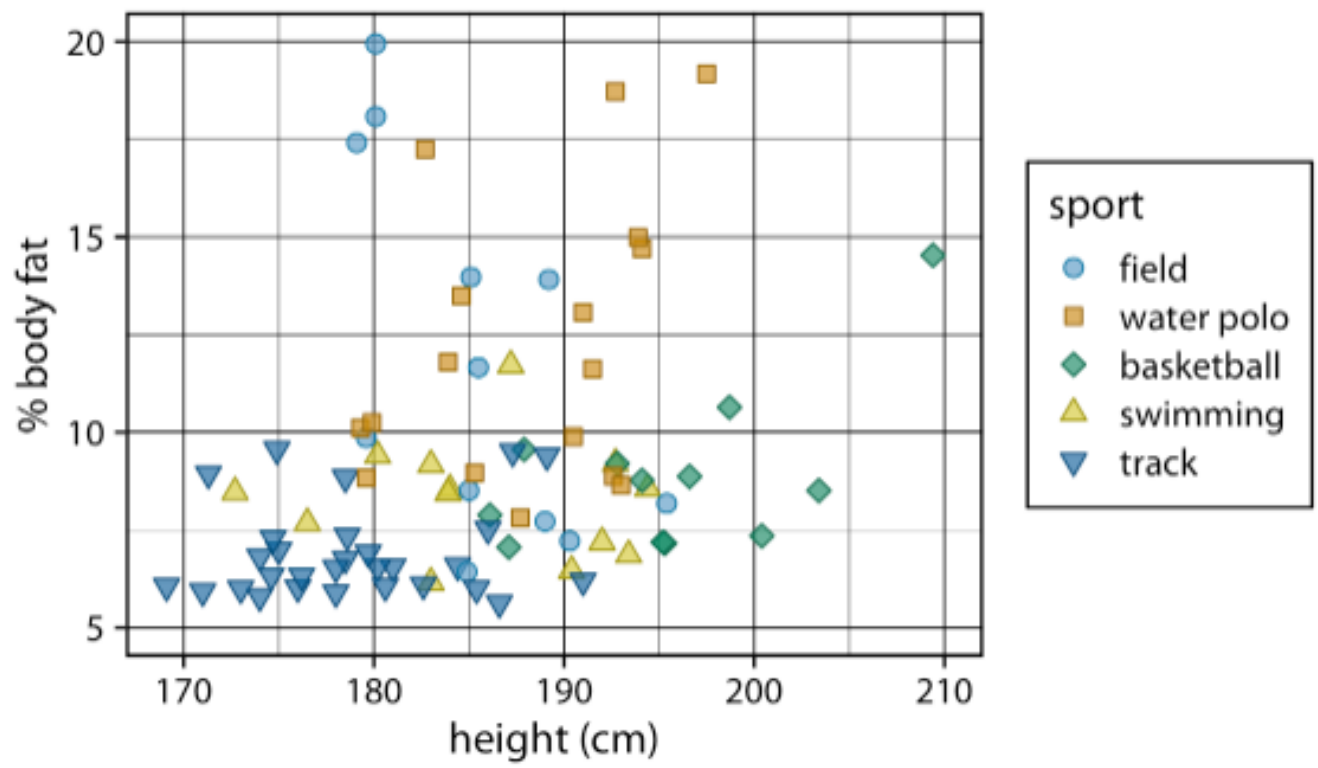
2) (5 points)

CASH CRUNCH
Net farm income in billions of 1967 dollars

3) (5 points)

We need national leadership...like Canada.
Covid is coded in red.

4) (5 points)

WILD FIRE FACTS
Fires caused by lightning vs humans in the US, 2019

Nº FIRES — ⚡ Lightning ☺ Humans — ACRES BURNT

Region	⚡ Lightning	☺ Humans
Alaska	371	349
Northwest	1430	2260
Northern California	420	3284
Southern California	76	4556
Northern Rockies	686	1623
Great Basin	921	1387
Southwest	786	1895
Rocky Mountains	736	948
Eastern Area	38	5712
Southern Area	898	22101

Acres burnt:
2,454,098
356,188
332,800
258,087
202,918
190,426
142,737
140,321
126,584
109,155
<100,000 EACH

@statistic.ly | Source: National Interagency Fire Center - nifc.gov/fireInfo/fireInfo_stats_lightng-human

.ıIstatistic.ly*

5) (5 points)

6) (5 points)

Relationship between food and health

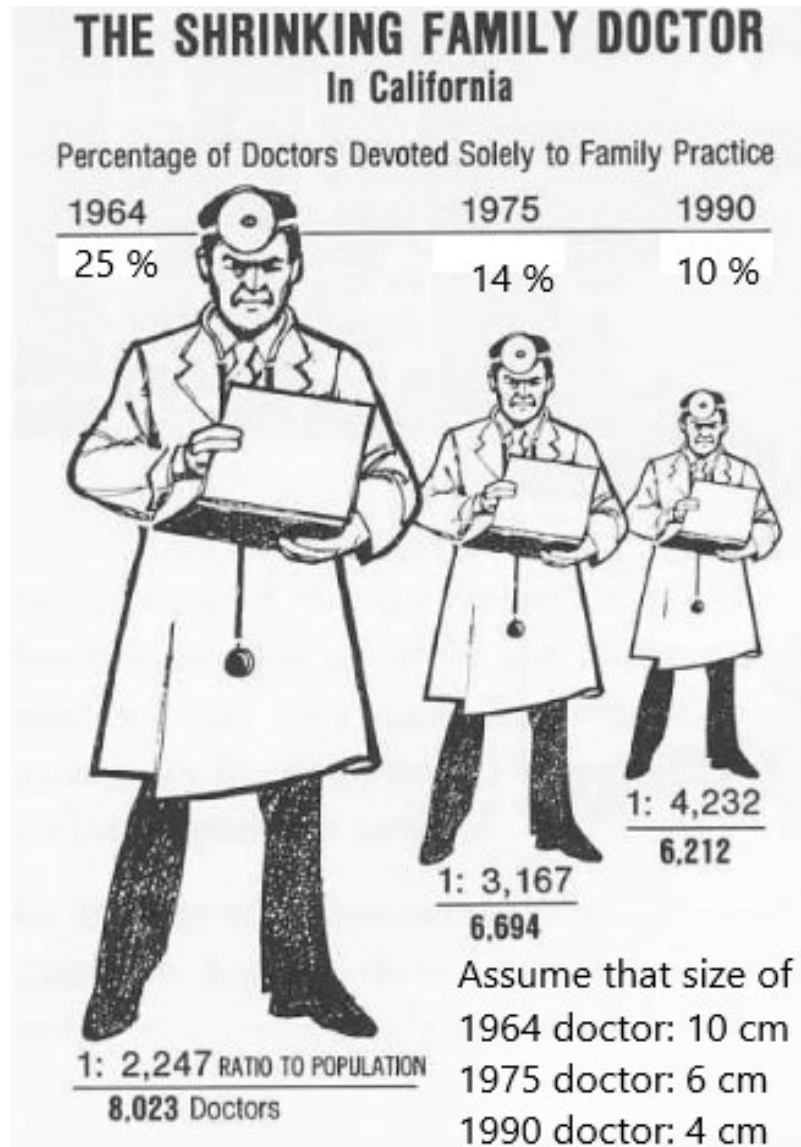**Question 2. (20 points)**

a) What are Edward Tufte's principles of Data Visualization? Explain them in your own words. (8 points)

b) Calculate the lie factor in 2 graphs below graphs. Please show your work and calculation. (12 points)

1) (6 points)

 *** Assume that 1964 doctor is 10 cm, 1975 doctor is 6 cm and 1990 doctor is 4 cm.

THE SHRINKING FAMILY DOCTOR
In California

Percentage of Doctors Devoted Solely to Family Practice

1964 — 25 %
1975 — 14 %
1990 — 10 %

1: 2,247 RATIO TO POPULATION
8,023 Doctors

1: 3,167
6,694

1: 4,232
6,212

Assume that size of
1964 doctor: 10 cm
1975 doctor: 6 cm
1990 doctor: 4 cm

2) (6 points)*Assume that the lie factor in Chart A=1; what could be the lie factor in Chart B?

Question 3. (40 points)

a) (5 points) What does this graph show? Who made it? When? What era of History of Data Visualization was it made?

b) (5 points) Name 5 things that this graph shows.

c) (30 points) Copy the master:

Using pandas and data visualization libraries in Python (matplotlib, Seaborn or plotly), replicate this graph. Provide your code in .py or .ipynb format and submit your code in pdf file.

Please follow the steps below:

In MyCourses under 3 text files containing the dataset for the above graph are provided: troops.txt, temps.txt and cities.txt. Draw a figure including 2 graphs using the data.

Step 1: Open Datasets

- Load both datasets (troops data and temperatures data) in Python.
- Display the first five and last five rows of each dataset.
- Show descriptive statistics (mean, standard deviation, etc.) for both datasets.
- Display the number of rows and columns for each dataset.

Step 2: Draw the First Graph (Troops Movement)

- **Create two lines in the chart representing troop size over distance.**
- The first line is the **"advancing line,"** showing movement marked by **"A"** in the data. Make the line **red** and adjust its thickness based on the troop size.
- The second line is the **"retreating line,"** showing movement marked by **"R"** in the data. Make the line **black** and adjust its thickness based on the troop size.
- Use the reference code from **"line_thickness.py"** to ensure the line thickness scales properly according to the troop size.

Step 3: Draw the Second Graph (Temperatures)

- **Create a second line chart representing the temperatures.**
- Assign the appropriate variable for the x-axis (e.g., distance or time) and the y-axis (temperature values).

Step 4: Add Labels to Both Graphs

- For the first graph (troops movement), add city names as labels to the line chart using the code from "**add_labels.py**".
- For the second graph (temperatures), add date values as labels on the line chart.

Step 5: Combine the Graphs in a Facet Diagram

- Create a facet layout that stacks the two graphs.
- Place the troop movement graph at the top and the temperature graph at the bottom.
- Ensure both graphs are aligned properly, and labels are clearly visible.

Question 4. (10 points) Read Lin and Thorton, 2022. Explain what colors are popular in data visualization? What types of charts are more attractive? What is considered as beautiful graph? What does it mean to be "fooled by beautiful data"?