

CSSE 386 – Data Mining with Programming  
**Rose-Hulman** Institute of Technology

Worksheet 03

Name (Print): \_\_\_\_\_ Section: \_\_\_\_\_

1. Name at least 2 items that Visual analysis can show:

- 1.
- 2.

2. What visualization can help improve (select all that apply)

- code practice
- decision-making
- information sharing
- model accuracy

3. Provide term for each definition

Term	Definition
	One variable
	2 variables / 2-dimensions
	Multi-dimensional

4. When to use each type?

Plot name	Description
	Useful in making comparisons between groups
	Useful when analyzing the data distribution
	Useful to find the relationship between variables

5. What is 5-number summary and sketch the boxplot

6. Find Median, 1st, 3rd, MIN, MAX , (IQR)

25, 28, 29, 29, 30, 34, 35, 35, 37, 38

## 7. Relationships in scatterplots: linear or non-linear

Method	Purpose
	a second variable changes proportionally
	the change in the second variable is not proportional

## 8. Review in class Dos and Dents and provide the main issue for each and how to fix it

- 1.
- 2.
- 3.
- 4.

## 9. Color Blindness Simulator

1. Use the Simulator link (course schedule for today)
2. Upload and test various images or slides
3. Change various views. Did it affect your colors?

Provide your observations:

## 10. Color Brewer

1. Use the Brewer link (course schedule for today)
2. Describe for:
  - Sequential data:
  - Diverging data:
  - Qualitative data:

Provide your observations (see questions on slides):

## 11. Matplotlib Gallery

1. Use the Matplotlib gallery link (course schedule for today)
2. Select 2 graphs, create a notebook, and replicate them

Provide the name of the selected plots: