URL to view Results	[[Click Here](#)]

Response Summary:

Mine Worksheet

Goal: to identify patterns, extreme and subtle features about the data
Objectives: Students will identify basic descriptors for the data, and categorize the data according to the specifications from the Parse Worksheet
Outcomes: Three (3) specific questions to be answered using the data

1. Student Information *

First Name	Gabby
Last Name	Willard
Course ***(e.g. CGT 270-001)***	CGT 270-LC4
Term ***(e.g. F2019)***	Fall 2021

2. Email Address *
   gwillard@purdue.edu

3. Visualization Assignment *
   • Lab Assignment

Analyze

4. Basic Descriptors: for each data component from the Parse Worksheet, identify basic descriptors (basic statistics). Explain *
   The maximum for age is 80, minimum is 0.17, mode is 24, median is 28, and average is 29.88. The most common age for the passengers is 24, the age that was in the very middle of the list is 28, the oldest passenger was 80, and the youngest passenger was 0.17 years old.
   The maximum for sibsp (number of siblings and spouses onboard) is 8, minimum is 0, mode is 0, median is 0, and average is 0.59. The most common number of siblings and spouses aboard is 0 and the highest number for siblings and spouses onboard is 8.
   The maximum for parch (number of parents and children onboard) is 9, minimum is 0, mode is 0, median is 0, and average is 0.39. The most common number for parents and

children aboard is 0. The highest number for parents/children is 9.
The maximum for ticket is 3101298, minimum is 680, mode is 1601, median is 234604, and average is 249039.1. The highest ticket number is 3101298, the lowest is 1601, and the ticket number that is in the middle of the list from lowest to highest is 234604.
The maximum for fare is 512.33, minimum is 0, mode is 8.05, median is 14.25, and average is 33.35. The maximum fare number is 512.33, the lowest is 0, the fare number in the middle of the list from lowest to greatest is 14.25, and the most common fare number is 8.05.
The maximum for boat is 16, minimum is 1, mode is 13, median is 10, and average is 9.5. The highest boat number is 16, the lowest boat number is 1, the boat number in the middle of the list from least to greatest is 10, and the most common boat number is 13.
The maximum for body is 328, minimum is 1, median is 155, and average is 160.8. The highest body number is 328, the lowest body number is 1, and the body number in the middle of the list from least to greatest is 155.
The mode for pclass (passenger class) is 3. The most common passenger class was 3rd class.
The mode for survived is 0. The most common survived number (from the number 0 to 1) is 0.

5. Categorize: consider what is similar and what is different? Categorize the data. Are the variables categorical (normal, ordinal, or rank). Are they quantitative (discrete or continuous)? Show categories. Explain. *

The nominal variables are survived, sex, and embarked. Survived can only be 0 or 1 with one of them meaning survived and the other meaning didn't survive, sex is either male or female (according to the dataset), and embarked is either the character, S or C.
The ordinal variable is pclass (passenger class). Passenger class is ordinal because it is a ranking system from 1 to 3 with 1 being the highest (best) class.
The interval variable is ticket. The ticket numbers do not start at 0, instead they start at 680.
The ratio variables are age, sibsp (number of siblings and spouses onboard), parch (number of parents and children onboard), fare, boat, and body. All of these variables have a zero point where calculations can be performed on all of them.

6. Temporal: is the data streaming data? How is it stored (all at one time, over several years in years, days, minutes, seconds)? Explain. *

The data has already been stored all at one time years ago because the tragedy of the Titanic happened a long time ago and is not a repeating event.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *

The data is mostly evenly spread because the values vary on both sides of the means from the variables.

Evaluate

8. Questions and Assumptions: list at least 3 questions you plan to answer with the data or list the questions if they were provided. Must be complete sentences and end in a question mark. What assumptions are you making? *

Question 1	How many people are numbered in the dataset?
Question 2	Which is the location that was the most embarked from?
Question 3	What country had the most people on the Titanic?
Assumptions	I am not aware yet of what the S and C stand for in the embarked variable. I am still not sure what the meaning behind the values 0 and 1 for the survived variable. I'm still not fully certain of what the numbers in the body variable mean.