

[Download as PDF](#)

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 first variable in the dataset was not given a name, so for this worksheet, I am going to call it "number." The maximum for number is 386, the minimum is 0, the mode is 0, the median is 136, and the average is 149.4191. That means that the largest number is 386, the smallest number is 0, the most common number is 0, the number in the middle of the ascending list is rounded to 136, and the average number is calculated at 149.4191.

The maximum for age is 66, the minimum is 15, the mode is 32, the median is 30, and the average is 30.69892. That means that the oldest crew member was 66, the youngest was 15, the most common age amongst the crew members was 32, the age in the middle of the ascending list is 30, and the calculated average age was 30.69892.

The mode for Survived is 1 according to my dataset calculation, but this could be an assumption because the original author of the data had started to enter zeros and ones in for this variable, but stopped early on in the column and left the boxes that were supposed to have zeros blank and only put in the ones.

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 Crew_Type, Embarked, and Survived because the values of these variables do not need to be ordered. These variables are all discrete because they are not changing over time.

The ordinal variable is Position because it is categorized in a specific order. This variable is discrete because the data is not going to change.

The ratio variables are number and Age because they all start at a zero point. This variable is discrete because the numbers and ages are displaying the information from the event of the Titanic crash.

There were no interval variables.

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

This dataset is not streaming data. It was stored all at one time recording information from the Titanic which is an event in history and is not changing.

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

The distribution of the data is mostly evenly spread. There are a few boxes that are either blank or have question marks. But other than that, the data is mostly distributed reasonably along both sides of the means on both sides of all of 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 sure of what the zeros and ones in the Survived variable represent. I am also not sure of what the numbers inside the first column mean because that column was never given a variable name and the numbers don't just count, they repeat.
