Names: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Plots in Minitab: LEGO Activity

In this activity, you will practice choosing plots that best represent relationships between variables and creating those plots in Minitab. After reading each question, select and create the plot out of these options that best visualizes the data:

scatterplot • stacked dot plot • histogram • box plot • pie chart • bar chart

There may be more than one correct plot for each question, but to complete the activity, you may only choose each plot type once.

The data you will be using was downloaded from a data collection platform called Kaggle and was modified by one of last year’s data science classes. It includes the following variables (you will only be working with some of them):

* **set\_id:** the unique identification number of each set.
* **set\_name:** the name of the LEGO set.
* **year\_released:** the year the LEGO set became publicly available for purchase.
* **num\_of\_pieces:** the piece count of the LEGO set.
* **set\_theme:** the theme of each set (examples include LEGO City, LEGO Friends, etc.).
* **audience:** the gender the set was marketed towards (girls, boys, or neutral).
* **brand\_partnership:** whether the set was made in collaboration with another company or brand (examples include Disney, Star Wars, Harry Potter, etc.).
* **set\_type:** the “type” of the set (toy or collector’s item).
* **name\_length:** the number of characters (letters, numbers, punctuation, spaces, etc.) in the name of the LEGO set (for example, the word “dog” has 3 characters; the phrase “my dog” has 6)

Instructions for how to make each plot in Minitab are below. If you need more help, just ask. ☺

**Scatterplot:** graph -> scatterplot -> simple, input X and Y variables, click OK.

**Stacked dot plot:** graph -> dotplot -> simple, input Y variable, click OK.

**Histogram:** graph -> histogram -> simple, input Y variable, click OK.

**Box plot:** graph -> boxplot -> simple OR with categorical variables, input Y variables (and categorical, if applicable), click OK.

**Bar chart:** graph -> bar chart -> counts of unique values -> input variable, choose categorical variable, click OK.

**Pie chart:** graph -> pie chart -> counts of unique values, input categorical variable, click OK.

1. We want to examine the distribution of the number of pieces included in LEGO sets. Create an appropriate visualization below.
2. How would you describe the distribution of this data?
3. Why did you choose this visualization?
4. Does there appear to be a relationship between the number of pieces included in a LEGO set and the year the set was released? Create a visualization that helps answer this question.
5. How would you describe the relationship? If you observe no relationship, how can you tell?
6. Why did you choose this visualization?
7. Out of all the LEGO sets in our sample, what gender (boys, girls, or neutral) was the target audience of the most LEGO sets? Create a visualization that helps answer this question.
8. The correct visualization should show you all three genders. Which gender had the smallest number of sets?
9. Why did you choose this visualization?
10. Does there appear to be a relationship between set type and the number of pieces in the set? Create a visualization that helps answer this question.
11. According to your visualization, which type, toy or collector’s item, has more outliers?
12. Why did you choose this visualization?
13. We want to examine the distribution of name lengths of LEGO sets. Create an appropriate visualization.
14. How would you describe the distribution of this data?
15. Why did you choose this visualization?
16. Are there more LEGO sets that were created with brand partnerships or without? Create a visualization that helps answer this question.
17. Approximately how many sets involve brand partnerships? How many do not?
18. Why did you choose this visualization?