

1. Open your file vgsales\_clean.xlsx in Excel.
2. Create a box and whisker plot of NA\_Sales. Describe it in your “Answers” sheet:
  - What are the Q1 and Q3 values?
  - Is the mean higher than the median or vice versa?
  - What does this tell you about the distribution of the variable?
  - Why do you think the plot looks the way it does?
3. Filter NA\_Sales to include only values less than or equal to 1 and create a new box and whisker of this filtered version. Describe the plot in your “Answers” sheet:
  - What are the Q1 and Q3 values?
  - Is the mean higher than the median or vice versa?
  - What does this tell you about the distribution of the variable?
  - Why do you think the filtering changed the plot?
4. Remove the filter on NA\_Sales. Create a 100% stacked bar chart with Genre on the x-axis and each of NA\_sales, EU\_sales, JP\_sales and Other\_sales making up the proportions (hint: you might try using a pivot table to do this).
  - Do any genres have a particularly high proportion of Japanese sales?
  - Which genre has the highest proportion of North American Sales?
  - Write your responses in the “Answers” sheet.
5. Filter the data to include only games from the shooter genre and create a scatterplot of NA\_Sales and Global\_sales.
  - How would you describe the relationship between these two variables?
  - Is this surprising to you? Why do you think the data suggests this relationship between these variables?
  - Notice that this plot looks pretty compressed. Why do you think the visualization looks this way? How might you classify that very top dot?
  - Record these observations in your “Answers” sheet.