Excel Workshop for GT SHPE

This workshop covers some basics of data formatting, visualization and brief mentioning of analysis. To follow through this workshop, we provided two sample datasets. Please download all content before starting.
The exercises are designed using Excel 2016. Please be aware of possible differences if you are using another version of Excel.

Data Formatting:

1. **Delete all blank rows:**
   - Select a column: Edit -> Find -> Go To – Special -> Blank -> OK
   - Delete all blank selections: Edit -> Delete -> Rows

2. **Fill empty cells in column "Serial Number":**
   - =IF(D2<>"","null","")

3. **Fill down in all empty cells in column "Manufacturer":**
   - Select all blank rows: Edit -> Find -> Go To – Special -> Blank -> OK
   - The first empty cell is auto-selected. Go up to formula: = the tell above it, cmd + enter for Window and cmd+return for Mac

4. **Remove trimming / leading white space in column "Accessary Title":**
   =TRIM(text)

5. **Concatenate cells:**
   =CONCATENATE(cell1," deliminator",cell2)

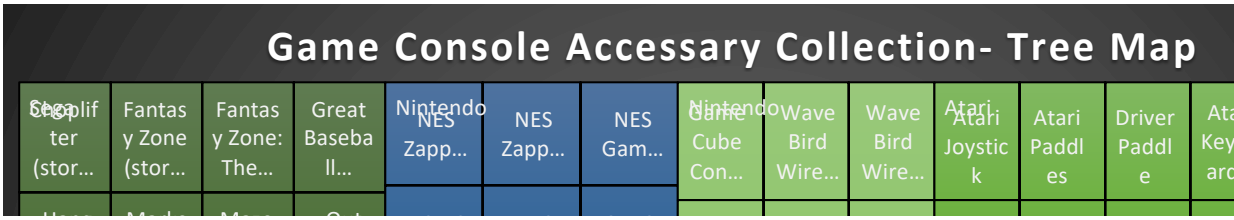
6. **Clean the column "Copyright Date"**

7. **Pivot Table:**
   - Pivot Table gives us an overview of …

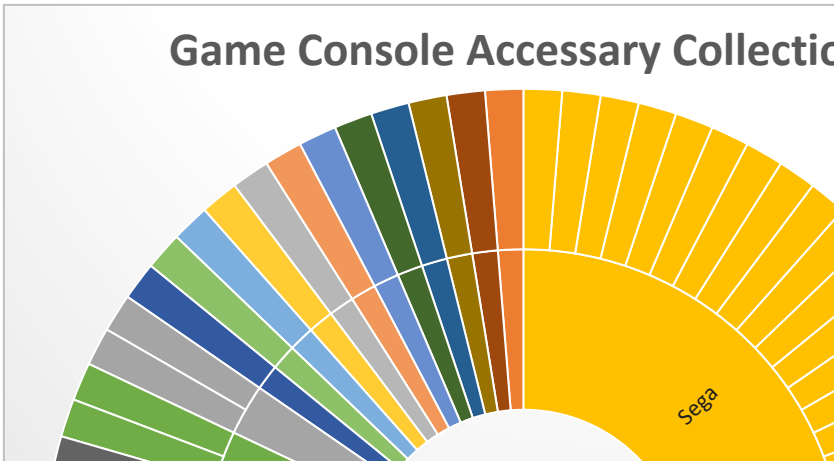
8. **Transpose data:**
   - Copy the whole table
   - Click on the first cell where you want to paste the table
   - Under Home, click on the drop-down menu next to Paste, Transpose.

Visualization:

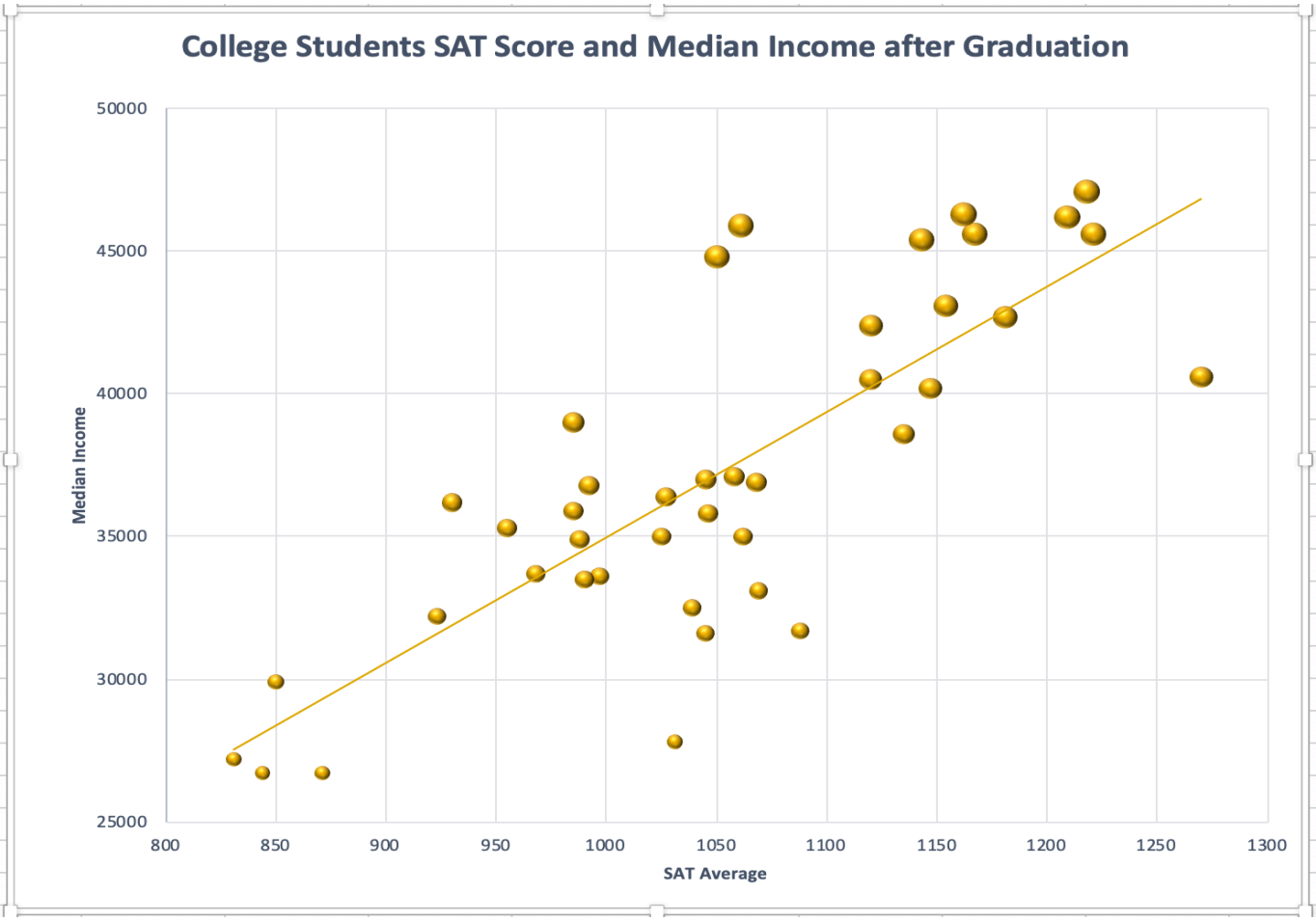
1. **Tree Map:** Manufacturer, Accessary Title, Count



2. **Sunburst:** Manufacturer, Accessary Title, Count, Copyright Date



3. **Bubble Chart:** SAT_AVG, MD_EARN
    a. Select the axis and change the range of data
    b. Select a bubble, right-click to Select Data, change size to MD_EARN
    c. Since the size difference is not significant enough, we could create a “Size” column to fill in value of the power of MD_EARN.



Data Analysis:

Data analysis functions are folded under Data -> Data Analysis. We will use sample data 2 to demonstrate two common analysis: correlation and regression.

1. Correlation Analysis

Select Data -> Data Analysis -> Correlation. Once selected a window would pop up, select the data range in the Input Range, check Label in First Row and an 4X4 area as output area.

The results for correlation is:

	UGDS_RICH	*SAT_AVG*	*MD_EARN*
UGDS_RICH	1		
SAT_AVG	0.7949566	1	
MD_EARN	0.59231447	0.79697412	1

2. Regression Analysis

Select Data -> Data Analysis -> Regression. Once selected a window would pop up, select the dependent variable in Y range and the independent variables in X range.

The results for regression is:

Regression Statistics	
Multiple R	0.799868722
R Square	0.639789967
Adjusted R Square	0.621317660
Standard Error	3591.628320
Observations	42

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	893570890	446785445.1	34.63508370	2.2534E-09
Residual	39	503091967	12899794.02		
Total	41	1396662857			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-12639.33	7616.6721	-1.659424409	0.10505093	-28045.503	2766.84356	-28045.503	2766.84356
UGDS_RICH	-2640.8333	3733.02894	-0.707423741	0.48350934	-10191.597	4909.93043	-10191.597	4909.93043
SAT_AVG	48.87987840	8.738999343	5.593303814	1.9009E-06	31.20358373	66.55617373	31.20358373	66.55617373