

# Making Sense of Data with Excel

## Part 1

Temple Analytics Challenge 2017

Mary Garcia Charumilind, Associate Director of Performance Analytics

# Three parts of analytics

1. Back end: data and databases
2. Analysis: statistics
3. Communication: visualization

# Which topics should we cover today?

## 1. Back end: data and databases

- Intro to Excel for beginners
- Tips and tricks for dealing with data
  - Navigation and shortcuts
  - Filters, slicers, and sorting
  - Joining data from different sheets (vlookup)
- Using BIG data
- Cleaning messy data
  - Correcting inconsistencies
  - Data that should not have been lumped into one column
    - Text to columns, unpivot

## 2. Analysis: statistics

- Pivot tables

## 3. Communication – visualization

- Graphing essentials

# Hello, Excel

- Do you have **Excel 2016**?
  - Figure out which version you have: <https://support.office.com/en-us/article/About-Office-What-version-of-Office-am-I-using-932788B8-A3CE-44BF-BB09-E334518B8B19?ui=en-US&rs=en-US&ad=US>
  - Excel 2016 can be downloaded from TUPortal: Look for the “Microsoft Download” link in the list on the left-hand side
    - **Really Big data?** Choose the 64-bit version and use a PC
      - Microsoft Excel 2016 for Mac doesn’t have “Get & Transform” feature. This is a new feature which only applies to Excel 2016. “New Query” and “Query Editor” are included in “Get & Transform”, so they are not available either
      - The TU TECH Center has PCs with Excel 2016  
<https://computerservices.temple.edu/tech-center-software>
      - Fox MIS lab in Alter 602 has powerful machines, but 32-bit Excel

# Hello, Excel

- Home -> Format -> **Autofit column width**
- **Getting around**
  - Ctrl or Cmd + End
  - Ctrl or Cmd + Home
  - Click A1, Ctrl or Cmd + down arrow (same with up, down, right, Page Up, Page Down)
- **Selecting data**
  - Click and drag the puffy plus
  - Shift + arrow key
  - Ctrl or Cmd + Shift + down arrow (same with up, down, right)
  - Ctrl or Cmd + a to select all
  - Ctrl + Shift + 8 to select the dataset I am in right now

# Hello, Excel

- Home -> Sort & Filter -> **Filter**
  - Click the menu at the top of a row to sort or filter
    - omit one, select all, unselect all, select a few
  - Select and copy filtered data, create a new sheet, and paste your filtered data as a new dataset
  - See also: **insert slicer** when columns are hard to keep track of
- View -> Freeze panes -> **Freeze top row**
- List of **distinct or unique values** in a column or table
  - Select data, copy, paste into new location, select the new data
  - Data -> Remove duplicates

# Cleaning messy data - Basic

- **Copy and paste**

- Ctrl or Cmd + c, Ctrl or Cmd + v
- Or use the skinny plus to drag and replace in any direction, or double click the bottom right-hand corner of the cell to be copied all the way down the column
- Right click or Cmd + click to copy, paste, or, paste special (values only, formats only, formulas only)

- **Find and replace** – use when a large proportion have the same typo

- Ctrl or Cmd + f
- Click to the replace tab
  - Find space and replace with [no space]

# Cleaning messy data - Intermediate

- **Trim** formula removes spaces before and after words
  - =trim(...)
- **Proper** formula turns all caps or no caps to Title Case
  - =proper(...)
- **Text to columns**
  - Make sure you have enough empty columns to the right
  - Select the column
  - Data -> Text to columns
    - Delimited (usually)
    - Select delimiter(s) or type one in there
    - Click Next, Finish

# Cleaning messy data - Advanced\*

- **Dealing with a list of things in each cell**

- If you've used Query Editor and changed the default data load settings before, do this:
  - Data -> Get Data -> Launch Query Editor
  - File-> Options and Settings -> Query options -> default query load settings.
  - Close the query editor window to go back to your sheet
- **Copy and paste** the problematic column(s) and IDs to a new worksheet and **Unpivot:**
  - Copy and paste just the ID column and the columns you want to unpivot
  - Select all, Ctrl + t to convert to a table
  - Data -> Get data from table/range (opens Query Editor)
  - Select ONLY all columns you want to unpivot (click column heading, Shift + arrow key)
  - Transform -> unpivot
  - Close and Load to your worksheet
- **Insert Pivot table:** IDs as rows, values as columns, attribute count as values
- Copy and **paste values into a new sheet.**

\*Microsoft Excel 2016 for Mac doesn't have "Get & Transform" feature (Query Editor). This is a new feature which only applies to Excel 2016.

# Combining datasets (Vlookup)

- Sort the ID ascending in your sheets!
- =vlookup(ID cell in destination sheet, click top of columns of the data to search, number of column the answer is in, FALSE)
- Copy the formula down (double click skinny plus on the bottom right corner)
- Paste values of your formulas
- (Repeat as needed)

# Common Problems with Vlookups

- The range reference was not anchored or did not use whole column
- The data wasn't in the column I thought
- The ID had a space at the end in one sheet but not the other
- The ID formats did not match (number stored as text in one of the sheets, etc.)
- The formula was left in (values weren't pasted), and then things moved

# Pivot Tables

- Click a cell in the table; Insert -> Pivot Table
  - Click Ok
- “To build a report” field list; drag them into the boxes
  - Rows: Use multiple fields and drag to reorder to show hierarchy
  - Columns
  - “Σ Values”: Click the drop down arrow inside the box and select Value Field Settings Average to change from the default
    - Click the Show Values As tab to Show as %, index, etc.
  - When you click out of the pivot table, the field list disappears.
- Copy, paste (or paste values), and graph
- The Analyze tab
  - Refresh (data)
  - Resurrect field list.

# Appending huge datasets

- Open Excel
- Data -> Get data -> Query options
  - default query load settings (under Global -> Data Load)
    - Specify custom
      - Load to data model only
      - Click ok
- Data -> Get data -> Launch **Query Editor**
  - New source -> file -> Excel
    - [repeat for each file]
  - Append queries -> Append queries as new
    - Select “Three or more tables”, add all
- Close & Load -> Close and load to -> only create connection , check “add to data model”
- Use Power Pivot add-in for Office 2016 Pro
  - File -> options -> advanced -> under Data, check “Enable Data Analysis add-ins: Power Pivot etc.”
  - In Power pivot tab -> manage
    - See your query as a sheet tab at bottom
    - Click Pivot table in the ribbon and add it to your worksheet.
    - Save from the Power Pivot window (click manage again) to avoid crashing

# Questions?

[mgcharumilind@temple.edu](mailto:mgcharumilind@temple.edu)

215-204-3208

Office hours: Mon & Tues 12pm-1pm through October 17<sup>th</sup>, Alter 731