

The Data Analysis Process

1 Import and Merge Datasets

Download data and combine datasets into a single master spreadsheet to use for analysis.

Text to Columns | vlookup | hlookup

2 Navigate Spreadsheet

Insert new sheets, freeze panes, and add filters to stay organized.

Sheets | Panes | Filters | Excel Tables

3 Clean Data

Check for duplicates, check for missing data, and recode variables.

Highlighting Duplicate Values | Counta, Countblank, and Sum | If and Ifs

4 Explore Preliminary Patterns

Use instant color-coding techniques and spark lines to find at-a-glance patterns.

Data Bars | Color Scales | Top/Bottom Rules | Spark Lines and Spark Bars

5 Calculate Descriptive Statistics

Use formulas or pivot tables to find the mean, standard deviation, and other measures.

Min, Max | Mean, Median, Mode | Standard Deviation | Countif | Pivot Tables

6 Calculate Inferential Statistics

Find correlations and calculate parametric and nonparametric tests to assess significance.

SPSS, SAS, and Other Statistical Packages or Excel Plug-Ins

7 Communicate Findings

Summarize findings in tables and graphs. Share information in reports and presentations.

Data Visualization and Reporting Techniques

Step

Skills, Formulas, and Hands-On Practice

2 Navigate Spreadsheet

Insert new sheets, freeze panes, and add filters to stay organized.

Dataset:
Demographics

Organize Sheets

- Creating
- Copying
- Moving
- Re-Naming
- Color-Coding
- Password-Protecting
- Hiding and Unhiding

Freeze Panes

- Freeze First Row
- Freeze First Column
- Freeze First Row *and* First Column

Sort and Filter

- Option A: Data → Sort
- Option B: Home → Sort and Filter → Filter

Hands-On Practice #1

1. Create a copy of the “Demographics” tab.
2. Move “Demographics (2)” so that it comes after “Demographics.”
3. Re-name “Demographics (2)” as “Demographics – Clean.”
4. Change the color of “Demographics – Clean.”
5. Freeze the first row and first column of “Demographics – Clean.”
6. Add filters to the first row of the dataset.
7. Using the new filters, sort the “Age” column from Least to Greatest.

3 Clean Data

Check for duplicates, check for missing data, and recode variables.

Dataset:
Demographics - Clean

Highlight Duplicates

- Home → Conditional Formatting → Highlight Cells Rules → Duplicate Values

Check for Missing Data

- Tallies: =counta, =countblank, =sum
- Color-Coding Empty Cells: Home → Conditional Formatting → Highlight Cells Rules → More Rules → Blanks

Recode Variables with =if

- DC → Yes, Not DC → No
=IF(H2="DC","Yes","No")
- Income Over \$100,000 → Yes, Not Over \$100,000 → No
=IF(D2>100000,"Yes","No")

Hands-On Practice #2

1. Highlight the duplicate ID numbers.
2. Delete the duplicate.
3. Use =counta, =countblank, and =sum to tally how many cells from each of the columns are missing data.
4. Color-code all of the missing cells in red; Fill in the missing data if you're able to.
5. Insert a new column to the right of "State." Use =if to recode cells so that DC = yes and MD or VA = no.

4 Explore Preliminary Patterns

Use spark lines and instant color-coding techniques to find at-a-glance patterns.

Dataset:
Exploratory

Spark Lines and Spark Bars (Excel 2010 and higher)

- Insert → Sparkline or Spark Bar ("Spark Column")
- Design → Show → High Point, Low Point, First Point, Last Point
- Adjust Sparkline Color and Marker Color
- Relative or Absolute Axes: Design → Axis → Vertical Axis Minimum and Maximum Value Options → Same for All Sparklines

Hands-On Practice #3: Spark Bars and Spark Lines

With the Virginia data...

1. Insert spark bars and spark lines below each column.
2. Adjust the Marker Color so that the largest bin stands out in a darker color.
3. Adjust both the minimum and maximum vertical axis values so that axes are the "Same for All Sparklines."

Conditional Formatting

- Home → Conditional Formatting → Data Bars
- Home → Conditional Formatting → Color Scales
- Home → Conditional Formatting → Highlight Cells Rules
- Home → Conditional Formatting → Top/Bottom Formatting

Hands-On Practice #4: Conditional Formatting

1. With the Chicago data, use data bars to insert miniature bar charts.
2. With the Test Scores data, use color scales to create a heat table.

5 Calculate Descriptive Statistics (with Formulas)

Use formulas or pivot tables to find the mean, standard deviation, and other measures.

Descriptive Statistics for Interval or Ratio Data

- Range
 - Minimum Value: =min
 - Maximum Value: =max
- Measures of Central Tendency
 - Mean: =average
 - Median: =median
 - Mode: =mode
- Measures of Dispersion
 - Standard deviation: =stdev
 - Variance: =var

Datasets:
Demographics - Clean
(Age; Household Income);
Survey

Hands-On Practice #5: Demographics - Clean

1. In Columns C and D, calculate the minimum and maximum values.
2. In Columns C and D, calculate the mean, median, mode, and standard deviation.

Descriptive Statistics for Nominal or Ordinal Data

- **Frequencies:** =countif
e.g., in cell B21, type =COUNTIF(B\$2:B\$118,\$A121)
- **Valid Percents** via simple math calculations

Hands-On Practice #6: Survey

1. Using the =countif function, tally how many people responded Strongly Agree, Agree, Disagree, and Strongly Disagree for each survey item.
2. Using the =sum function, tally how many people responded to each survey item.

5 Calculate Descriptive Statistics

(with Pivot Tables)

Use formulas or pivot tables to find the mean, standard deviation, and other measures.

Datasets:
Demographics – Clean;
Technical Assistance Log

Insert a Pivot Table

- **Ensure that prerequisites are met**
 - Every column must be labeled
 - Contiguous cells (No completely empty rows or columns)
- **Insert a Pivot Table:** Click on the cell in the upper left-hand corner of your data table → Insert → Pivot Table
- **Drag and Drop Variables**

Hands-On Practice #7: Demographics - Clean

1. **Frequencies**
 - a. Number of males: _____
 - b. Number of females: _____
 - c. Number who work for the government: _____
 - d. Number who are retired: _____
2. **Crosstabs**
 - a. Number of males who are self-employed: _____
 - b. Number of females who work for the government: _____
 - c. Number of people with 15-20 years of experience who live in Fairfax: _____
3. **Filters**
 - a. Maryland only: Number of males: _____
 - b. Virginia only: Number of females: _____
4. **Explore missing data**
 - a. Number of people with no state data: _____
 - b. Double-click on mysterious entries; make corrections where possible

Refreshing Data

- As you're actively making changes: Click Refresh icon in pivot table tab OR right-click within pivot table and select Refresh)
- Automatically every time you open the file: Pivot Table Options → Refresh data when opening file

Hands-On Practice #8: Technical Assistance Log

1. Summary Statistics/Exploring the Dataset

- a. General date range: _____
- b. Number of technical assistance providers: _____
- c. Number of organizations receiving assistance: _____
- d. Types of assistance provided: _____
- e. Total amount of assistance provided (minutes) : _____

2. Crosstabs

- a. Amount of assistance provided by Carlos: _____
- b. Amount of assistance provided by Carlos to the Sycamore Food Bank: _____
- c. Amount of assistance provided by Carlos to the Sycamore Food Bank in September: _____
- d. Organization that received the most assistance (minutes) : _____
- e. Organization that received the most assistance (minutes) related to IT systems: _____
- f. Person who provided the most data and evaluation assistance: _____
- g. Person who provided the most data and evaluation assistance in October: _____

3. Grouping

- a. Grouping by month: Minutes of assistance provided in October: _____
- b. Grouping by day: Minutes of assistance provided on May 18: _____
- c. Grouping by week: Minutes of assistance provided between 1/3/2014 and 1/9/2014: _____
- d. Alternative grouping strategy: Create new variable using =month() function

Additional Time-Saving Techniques

Dataset:
Time-Saving Techniques

Dealing with Dates

- =month
- =day
- =year
- =concatenate
- &

Hands-On Practice #9: Dates

1. Find each date's month, day, and year.
2. Use =concatenate and & to combine the year and month into Year-Month format.

Dealing with Names/Text

- Text to Columns: Highlight the cells that need to be separated → Data → Text to Columns → Delimited → Select the delimiter (e.g., a space) → Finish
- =left, =right, =mid
- =lower, =upper, =proper

Hands-On Practice #10: Names

1. Use "Text to Columns" to separate "ann emery" into "ann" and "emery."
2. Use =left to find the first letter of each person's last name.
3. Use =proper to transform "ann emery" into "Ann Emery."