DATA ANALYTICS: DYNAMIC DATA REFERENCING

INTRO VLOOKUP/HLOOKUP

VLOOKUP, HLOOKUP

- Easy and effective lookup functions that scan rows and columns for the desired data
  - They are better than normal searches because they are faster easier to change the query

- V-Lookups (vertical lookups) scan columns
- H-lookups (horizontal lookups) scan rows

VLOOKUP, HLOOKUP

- Most often used to perform quick searches for a single object in a long column/row

- **Function:** To run a search on a datatable by a value from the first column
- **Output:** A value from any cell on the same row from the first column value
- **Importance**: VLOOKUP and HLOOKUP allow you to connect and lookup information from different datasets.

=V/HLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])

Example of using VLOOKUP, HLOOKUP

- Imagine that you're a customer service rep needing to find how many purchases a customer had, using a VLOOKUP we can do this very quickly. You can create a second sheet and use LOOKUPs across the sheets to bring elements from the first sheet into the other.

UserID	User_Name	Profile_Creation_Date	Quick_Pay	Newsletter	Number_of_Previous_Purchases
100000	ecyw	11/5/14	0	0	7
100001	mihz5073	7/3/14	0	1	5
100002	khljhj3457	6/30/14	1	1	10
100003	bpax9421	8/7/13	1	1	9
100004	wdnms2654	8/15/14	0	0	3
100005	zvtr	6/25/14	0	0	9
100006	thk4855	5/15/14	0	1	5
100007	ugde142	6/26/14	0	1	1
100008	drwr8182	8/4/13	0	1	1
100009	gfp3531	1/30/13	1	1	10
100010	smv5083	6/28/13	1	0	10
100011	haive8272	11/29/14	0	1	5
100012	sckl1495	5/28/13	1	0	6

Practicing VLOOKUP, HLOOKUP

- We're going to be writing more and more powerful formulas as we progress in this class
- The goal, as an efficient analyst, is to decrease the amount of work you're doing - your work should be agile. If you see a task that is repeatable, you should aim to write a powerful formula to do the work for you in the future.

Practicing VLOOKUP, HLOOKUP

- Open the AN_Lesson_4_lookup_exercise_student.xlsx to the assignment tab. Our goal is to provide our hedge fund analysts with a summary tab of information regarding the security ticker, name, address, and GICS sector. With the instructor, fill in this tab using VLOOKUP for the top table and HLOOKUP for the bottom table.

- If this tab is filled in properly, we should be able to enter any ticker symbol and populate the remainder of the row.

Intro PIVOT TABLES

What are pivot tables?

- Aggregation technique for clean data that allows us to chart and visualize certain combinations of data.
- A good tool to find relationships between different columns - and subsets of those columns - in a table.
- Classifies numeric data in a list based on other fields in the list.

What can we do with pivot tables?

- Quickly summarize data from a worksheet or from an external source - calculate totals, averages, counts, etc. based on any numeric fields in your table
- Provide the most straightforward way to explore different, multidimensional aggregations in Excel
- Generate charts from your pivot tables
- Find relationships between columns
- Allow users to explore more about the datasets they're working with

When should we create pivot tables?

- When we want a way to smart autofilter based on advanced criteria
- If we know how to filter/advanced filter, pivot tables can be used in conjunction to:
  - Explore and hide extraneous data using table views, Autofilter, and criterion range
  - Take larger clean data sets and display specific subsets of data and pivot on it

The situation

- The hedge fund you're assigned to is considering expanding into other countries. Again, the analysts you're working with have asked you to provide aggregated, summarized data on expansion of imports and exports of countries in the world. Given the World Trade Org (WTO) data, create two summary tables that highlights the breakdown of Imports & Exports by sector. The analysts want to be able to toggle by country and by sector.

Our plan

- Before diving into aggregation, take two minutes to identify, with a partner, the business needs, and what questions you might be able to answer.
- After we go over the important questions we'll be able to answer, we're going to create two summary charts using separate pivot tables that answer the questions and create the pivot tables in such a way that a user can explore them to get further results.
- Then we'll brainstorm what we could send in an email to our analysts to give them a recap.

Business questions we should be able to answer

Overall trend questions:

- Which countries with the highest import growth in the last 5 years?
- Which have the lowest (or negative) growth?
- Which import indicators have the highest volume and which have the highest growth over the time of the data?

Using toggle by country table:

- What were Clothing Exports In Germany in 2007?
- What were total Imports in the the US in 2009?

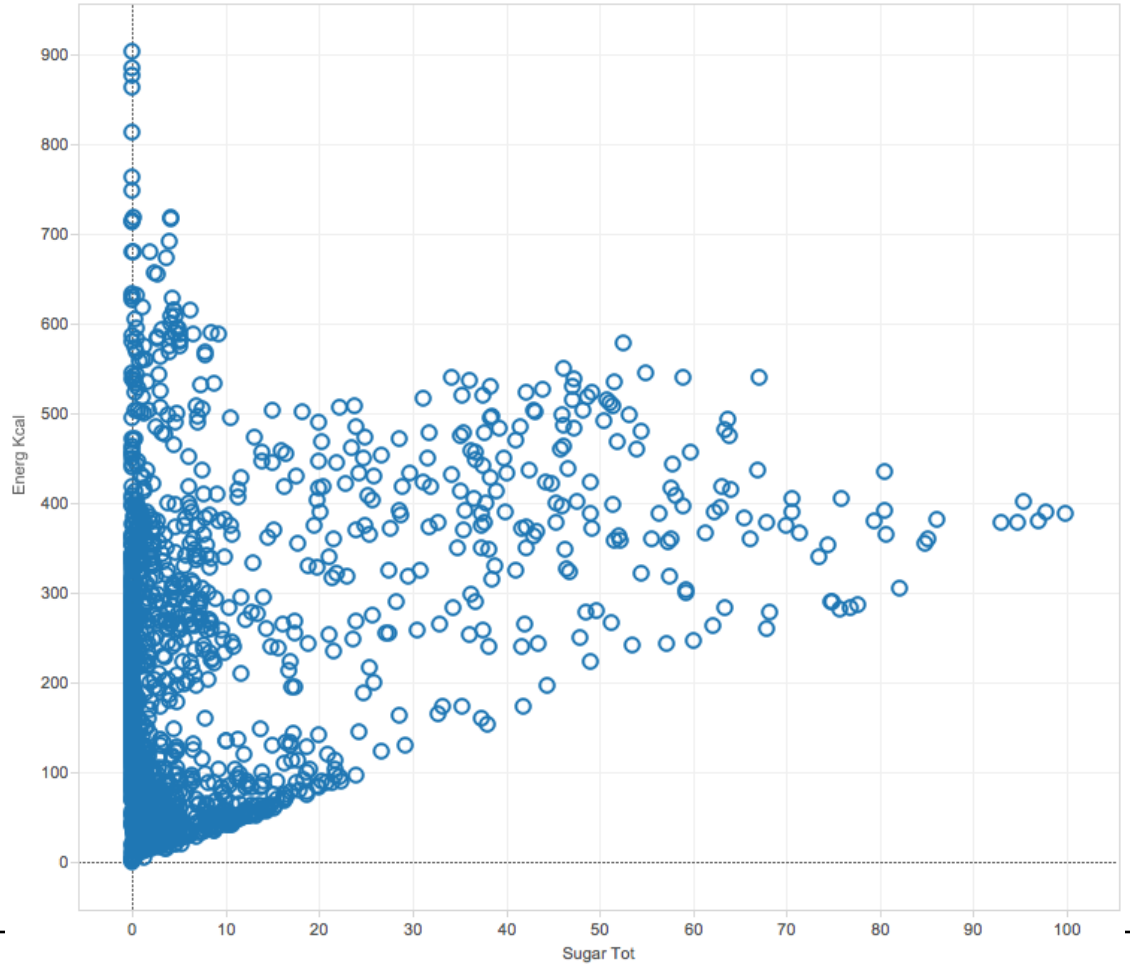
Using toggle by sector table:

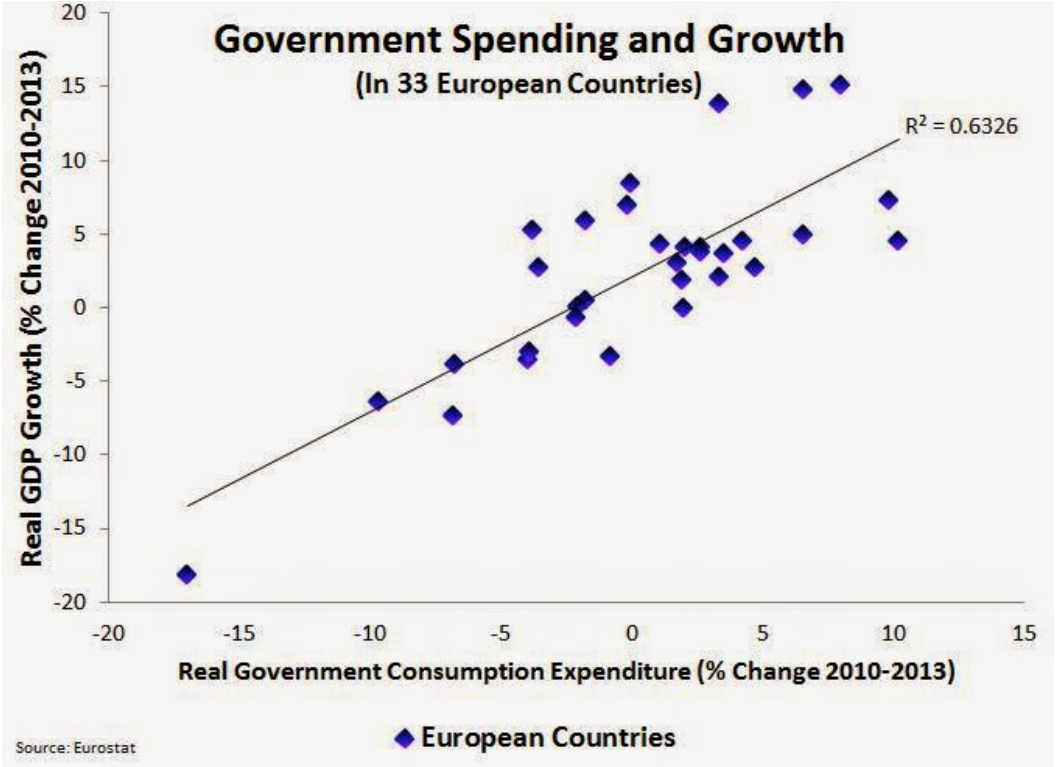
- Who were the three largest total exporters in the data in 2013?
- Who imported more fuel in 2011, the US or China?

INTRO SCATTER PLOTS

What is a scatter plot?

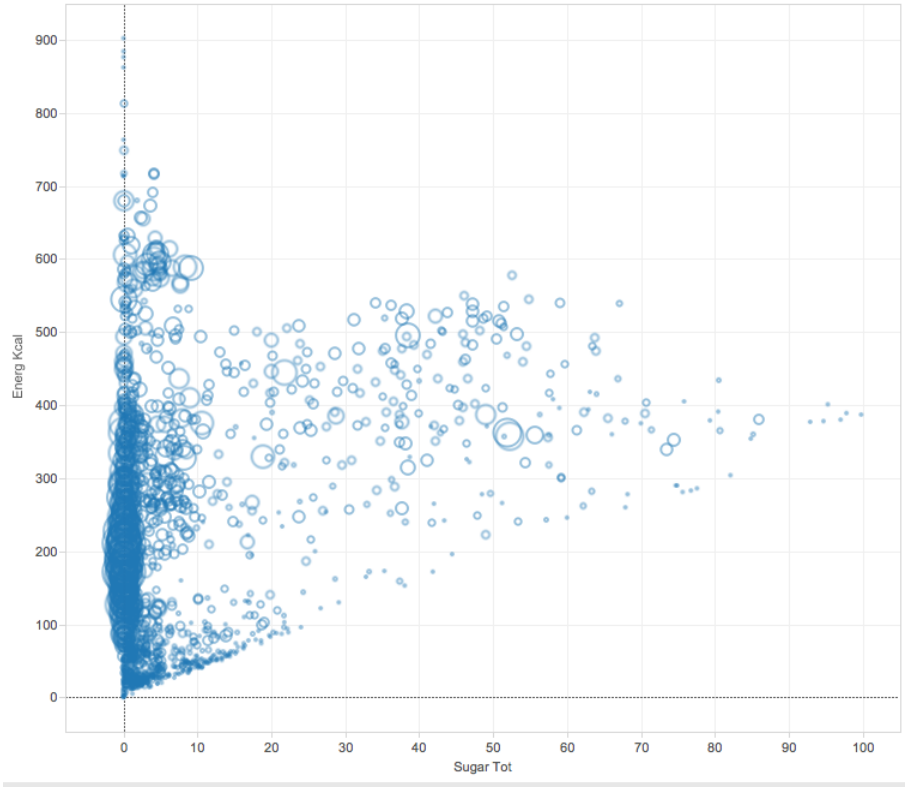
- A scatter plot is a chart obtained by mapping numeric values to a pair of orthogonal axes.
- Dots represent a data point in the data set and its position is given by the x and y position obtained by using the values of the two attributes that are mapped to the two scatter plot axes.


Energ Kcal
900
800
700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
70
80
90
100
Sugar Tot


Government Spending and Growth
(In 33 European Countries)
$R^2 = 0.6326$
Real GDP Growth (% Change 2010-2013)
Real Government Consumption Expenditure (% Change 2010-2013)
20
15
10
5
0
-5
-10
-15
-20
-20
-15
-10
-5
0
5
10
15
European Countries
Source: Eurostat

Scatter plot variants

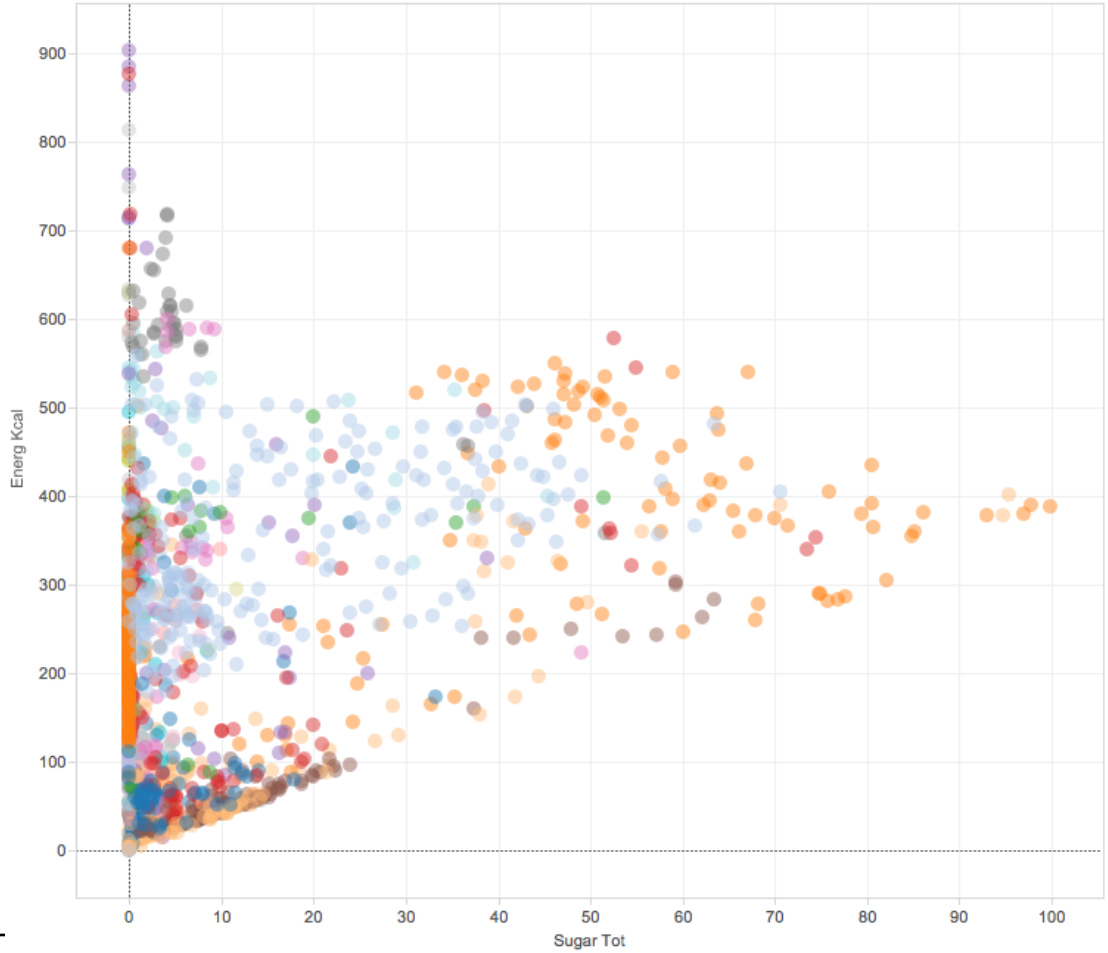
- **Bubble Chart**: same as scatter plot but a third numeric attribute is mapped to the size of the bubble.


900
800
700
600
500
400
300
200
100
0
Energ Kcal
0
10
20
30
40
50
60
70
80
90
100
Sugar Tot

Scatter plot variants

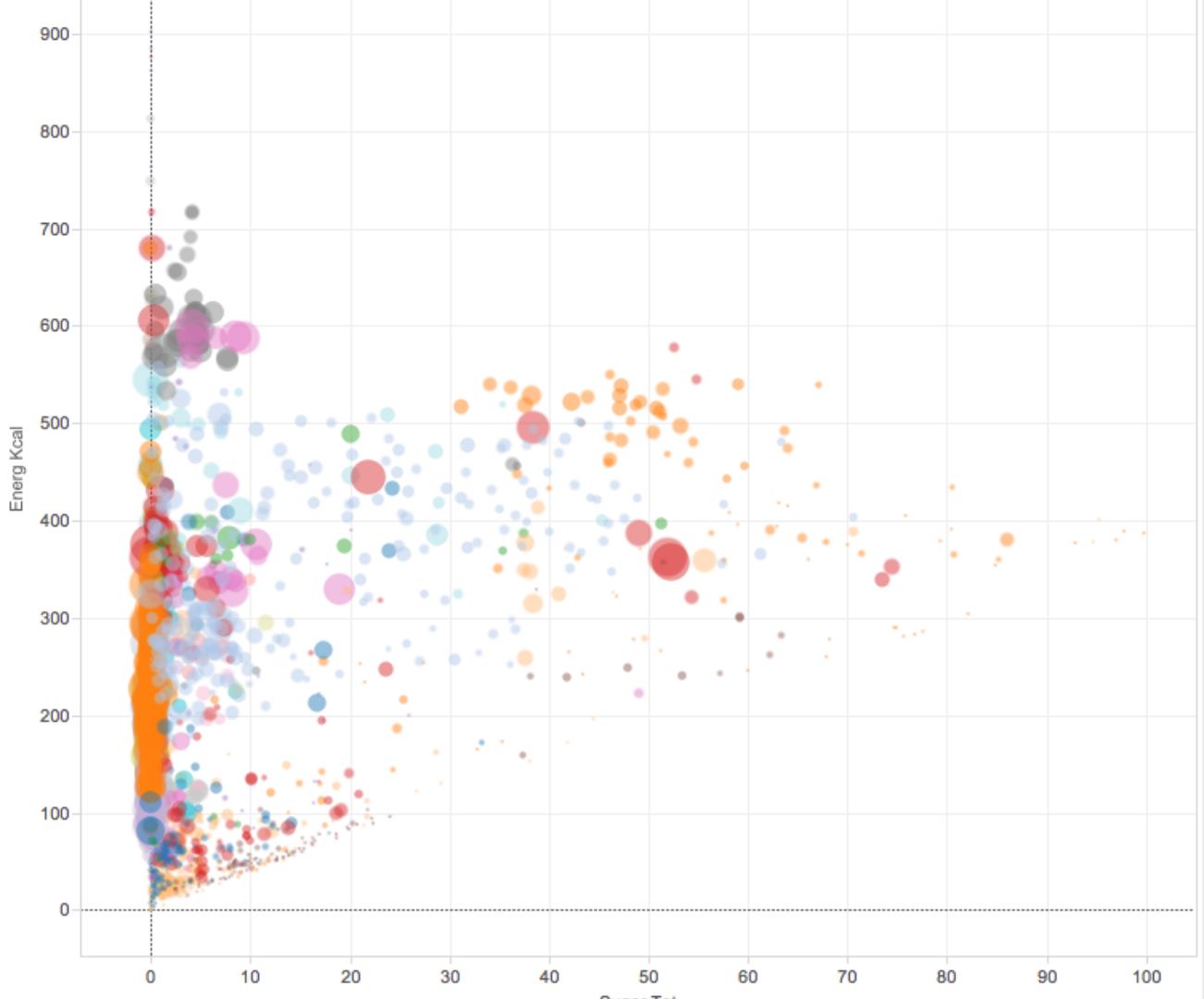
- **Scatter Plot with Colored Dots**: a scatter plot can also be enhanced by mapping the color of the dots to an additional attribute (the same way we just did with size). Color coding can be used both to map a categorical as well as a numeric attribute. Typically, categorical attributes work best on this case

Orange is sweet.



Scatter plot variants

- **A combination of bubble and colored dots:** In the figure we'll see, you can see how we, again, map size with color at the same time, effectively mapping a total of 4 attributes in one plot.


Energ Kcal
900
800
700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
70
80
90
100

Use cases and drawbacks

- Use case: to show the association between 2, 3 or 4 variables.
- Drawbacks:
  - The biggest issue with scatter plot is data size: if there are too many data points it can become too cluttered to show useful information.
  - Mapping color to an attribute with too many categories (hard to distinguish more than 5 or 6 colors).
  - Another similar problem is when the data is highly skewed the dots may all be concentrated in a small area.

Creating a scatterplot in Excel

- Looking to get more information, the analysts you're working with are now asking for information about the development of the countries you've gathered information; they are specifically looking to know more about the country's demographics.

- Given the World Health Organization data, create a scatter plot that demonstrates the relationship between Fertility & Life Expectancy globally.
- We'll be answering this question:
  - Look at three different years: 1960, 1980 & 2000. What do you think can be concluded, if anything given this data?

Colors

BLACK
--
CMYK 0/0/0/100
RGB 0/0/0
HEX 000000
PMS Process Black C

WHITE
--
CMYK 0/0/0/0
RGB 255/255/255
HEX FFFFFF
PMS -

RED
--
CMYK 0/100/100/0
RGB 229/27/36
HEX e51b24
PMS 485 C

YELLOW
--
CMYK 0/0/0/100
RGB 255/216/0
HEX ffdb00
PMS 108 C

MINT
--
CMYK 0/0/0/100
RGB 0/0/0
HEX 85e8da
PMS Process Black C

TEAL
--
CMYK 0/0/0/100
RGB 0/0/0
HEX 1ecac7
PMS Process Black C

BURGUNDY
--
CMYK 0/0/0/100
RGB 0/0/0
HEX 7a1743
PMS Process Black C

PINK
--
CMYK 0/0/0/100
RGB 0/0/0
HEX ffaec2
PMS Process Black C

LIGHT GREY
--
CMYK 0/0/0/10
RGB 29/29/29
HEX eaeaea
PMS 421 C

DARK GREY
--
CMYK 0/0/0/80
RGB 88/88/91
HEX 333333
PMS 418 C

INSERT TERM

Ipsum dolor sit amet...

Ipsum dolor sit amet...