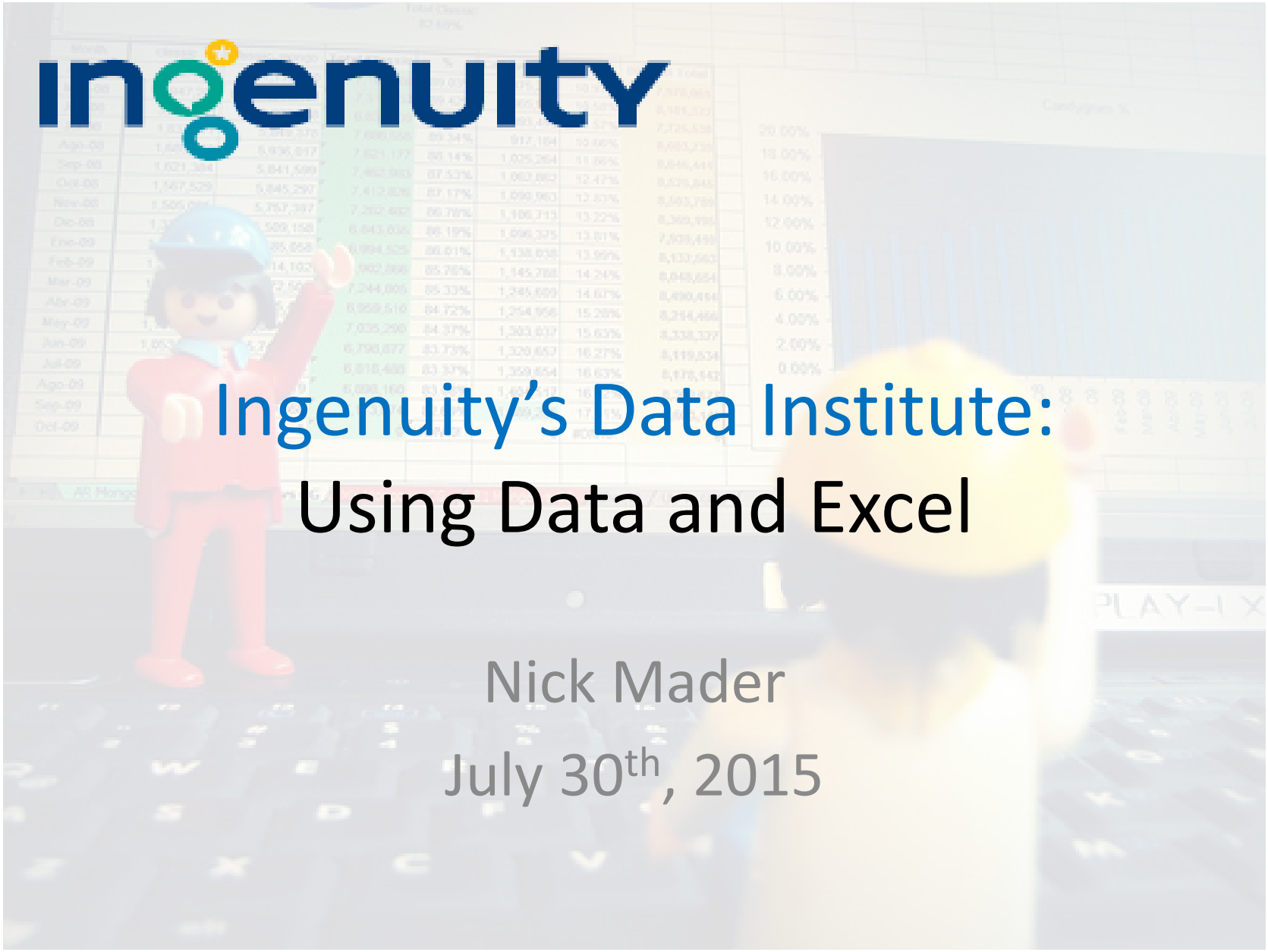




Ingenuity's Data Institute: Using Data and Excel

Nick Mader
July 30th, 2015



Excel

- Bob Morrison:

“It’s *not* rocket science!”

Welcome!

The plan for the next hour:

- Cooking show
 - Make something from scratch to finished product
 - Focus on some key concepts and ideas
 - Share every way I can think of to help you learn more beyond this single hour



Excel is so great

Things you know, since you're here:

- It's a common denominator—
 - You have it and know how to use it
 - Everyone you communicate with has it and knows how to use it
- It's got a (relatively) mild learning curve
- It's powerful
- It's flexible and transparent



Excel can be so much better

Things you may not know, since you're here:

- Being flexible, Excel won't stop you from bad design
- Working in spreadsheets can be tedious and time consuming
- When it feels like “there *has to be* a better way”, there *always* is



Principles we'll cover

- 1. Design/workflow** – function, clarity, efficiency
- 2. New tools** –
 - Lookup formulas
 - Relative vs absolute cell references
 - Range names
 - Filters and conditional formatting
 - Pivot tables
- 3. Learning more** – more time learning Excel means less time using Excel (in the good way)

Today

- For our demo, I've created (fake) data with baked-in challenges:
 - We are playing the role of a youth arts organization, looking at student school day attendance, by quarter, for all of the programs that we run
- This data, this presentation, the “answer key”, and other goodies from today's session will all be accessible at ...

Design/Workflow Principles

- Organize content
 - Design data tables so that the rows represent your key level of focus
 - E.g., if you're focused on students, have one row for each student, with all of their information going across in columns
 - Separate different content into different sheets
 - This allows for better clarity for users
 - It allows for better control on formatting, e.g. of row and column sizes
 - We'll look at nice ways to connect different sheets

Design/Workflow Principles

- Make your order of operations transparent
 - Never modify the raw data directly
 - Use multiple columns to show steps
 - Use colors to visually communicate to the reader
 - Use range names to make formulas readable
- Avoid work by hand
 - It's slow
 - It's prone to mistakes
 - It's not reusable
 - Tedious work takes time away from creative or reflective work

Design/Workflow Principles

- Make all your work flexible/general/reusable
 - Use (and seek out) formulas to do work that you would otherwise do by hand
 - Set things up once, so that (ideally) any updates will be effortless

Step 1. Centralize all of our data into one workbook

- Open all of the data files
- Get the data from each file into one workbook using **Home>>Format>>Move or Copy**
- Smarter, not harder: use **Alt + <key sequence>** instead of mouse to easily/rapidly perform operations



Sidebar: Navigating Shortcuts

- Learning shortcut keys can make your life in Excel much quicker and less tedious:
 - **Ctrl + Arrow Key** – Move to edge of data
 - **Shift + Arrow Key** – Expands the selection
 - **Ctrl + Shift + Arrow Key** – Move to edge of data while expanding selection

 - **Ctrl + Letter** – Short-cut for some functions
 - **Alt + Key Sequence** – Lets you access every single function
- Google (and [this great site](#)) have more shortcuts

Step 2. Establish a concrete goal for the work

- We'll want two tables: one where each student's record is the row, and one where each program is the row
- Make "table shells" which have the row and column labels to make clear what we're aiming for
- Create a new sheet "Student Data" to separate it from the raw data, and start it off with the data in "Program Assignments"
- Also use the tab color to set it off from the tabs that contain raw data

Step 3. Work attendance rates into our student table

- We need to go from daily attendance rates to quarterly attendance rates by youth
- Use **VLOOKUP()** to bring quarter data from the “School Calendar” sheet
 - Note the “on-board” help in the formula bar, which tells us what information the VLOOKUP() formula needs

Step 3. Work attendance rates into our student table

- Copy the formula all the way down
 - Double-click on the bottom right box of the cell selection; or
 - Use **Ctrl + arrow** and **Ctrl + Shift + arrow** navigation to copy the formula down
- Color code the text of the formula cells to set it off from the raw data

Step 3. Work attendance rates into our student table

- We'll use the **AVERAGEIF()** to calculate attendance rates by youth and quarter
 - We'll need to create a way to reference the student and quarter data in the attendance sheet (all in the row), and match that to student (in row) and quarter (in column) data in our final table
- Create a column in the "Attendance qtrs 1-3" sheet containing a combined student+quarter ID
 - In E2, enter **=B2&"_"&D2**
 - "&" is the "concatenation operator" – like an addition operator, except that it glues together text

Step 3. Work attendance rates into our student table

- Fill in the **AVERAGEIF()** formula to make the calculation
- Copy the formula to all student rows and attendance columns
- Oh no! It's busted! Why?!
 - Because Excel tries to be helpful and changes the cell references so that they have the same relative position to each cell that has that formula
 - In some cases we want this, in some cases we don't
 - Insert a "\$" to precede each part of the cell reference that we want to lock down
 - **=\$B3** could be copied left and right, but the column **B** won't change
 - **=B\$3** could be copied up and down, but the row **3** won't change

Step 3. Work attendance rates into our student table

- Clean up the formula
 - Remove unnecessary references to our own sheet, e.g. get rid of **'Student Data'**!
 - Name ranges so that they're more interpretable. Can do this either by:
 - Selecting the range and changing the name in the "Name Box" at the top left; or
 - Going through **Formulas>>Name Manager**
 - These names will be recognized (by autocomplete!) in Excel formulas

Step 4. Visually interpret data

- We'll use conditional formatting to help visually interpret data
- Apply it through **Home>>Conditional Formatting**
- Can use many different colors, icons, or rules for how to visualize higher or lower values

Step 5. Use Excel's filter to explore the data

- Apply it through **Data>>Filter**
- Allows:
 - Clicking the arrow shows you all unique values
 - Easy sorting → try sorting by Program
 - Filtering by categorical values → try filtering by program type
 - Filtering by numerical values → try filtering by a numeric range on one of the attendance fields
 - Sorting or filtering by color/categorizations from the conditional formatting

Step 6. Create a program-level table

- Create a new “Program Data” tab with its own table shell and tab color
 - To get a unique list of the program names, use **Data>>Remove Duplicates...** on the longer list of student assignments
- Connect this to the student-level data with another **AVERAGEIF()** formula

Step 7. Use a pivot table to summarize program data

- This works best for data that is as granular as possible, where rows represent each individual data point
- Start with the “Attendance qtrs 1-3” sheet and use **VLOOKUP()** to bring in program data

Step 7. Use a pivot table to summarize program data

- Create a pivot table through **Insert>>PivotTable**
 - Select the entire data range
 - In the wizard, select the “Program”, “Quarter”, and “Present” fields
 - Drag these to the Row, Column, and Values boxes respectively
 - Click on “Count of Present” in the values field and select **Value Field Settings...>>Average**
- Pivot table options can be controlled from special **Options** and **Design** menus at the top, which appear when you click on a pivot table

Step 7. Use a pivot table to summarize program data

- Pivot Table Pros
 - Is easier when doing complicated calculations
 - Great interactivity for designing tables
 - Can create sophisticated references with “Show Values As...” options (under “Value Field Settings...”)
- Pivot Table Cons
 - Values cannot be edited (good and bad)
 - Can get around this by copying and pasting elsewhere
 - Formulas referencing pivot tables work in an unexpected way
 - Tables do not automatically update (but *can* be manually refreshed)

Step 8. Add new data, and see how well our tables update

- Open the “Attendance qtr 4” file
- Copy and paste the data at the bottom of the “Attendance qtrs 1-3” sheet
 - (Note that our earlier range-naming and formulas extended across the entire columns, and thus will encompass all this new data)
- Extend the formulas to the bottom of the table

Step 8. Add new data, and see how well our tables update

- In each, copy the column for Quarter 3 over, and retitle it to “4”
- That’s it—we’re done
- Good design!

Step 8. Add new data, and see how well our tables update

- Right-click on the pivot table, and select “Refresh”
 - (Or we may need to go to **Options>>Change Data Source...** to update the range)
- That’s it, we’re done
- Good design!

Learning Principles

- Taking time to learn always pays off
 - We'll be working in Excel for the rest of our careers
- When it feels like “there *has to be* a better way”, there *always* is
 - Excel is so well developed, that someone has made a solution to your problem
- Your best solution may not always be in Excel

Learning Practices

- See the 5 part “Excel for non-profits” [webinar series](#) I did for Data Analysts for Social Good
- Take time for unstructured browsing of Excel’s functions
 - Notice unfamiliar commands in the menu bars
 - In the **Formulas** menu, look into all the formula groups (e.g. “Text” or “Lookup & Reference”)
 - Browse sections in <http://www.excel-easy.com/>
- Use “[pair \(or mob\) programming](#)” to help staff learn from each other
 - Good for learning both big features and nitty-gritty
 - Good for culture

Learning Practices

- Consider other (free, open source) hammers
 - [OpenRefine](#) – specifically designed to anticipate and help clean data
 - [R](#) – more learning curve, but allows faster/more sophisticated/lovelier work
- Come to Chi Hack Night – <http://chihacknight.org>
 - We meet every Tuesday at 6:00pm at Braintree in Merchandise.
 - It's free. Dinner is sponsored. There are both learning and project groups. It's diverse: age, type of expertise, beginners to advance dfolks.
- E-mail me!
 - nsmader@gmail.com