**Spreadsheet Best Practices/Run Charts Workshop January 2020**

**Companion worksheet**

The inspiration for this workshop is the data carpentry workshop on spreadsheets.<http://www.datacarpentry.org/spreadsheet-ecology-lesson/>

Resources: Headfirst Excel O'Reilly Books – <http://uclid.uc.edu/record=b6626353~S39>

Hadley Wickham – Tidy Data <https://www.jstatsoft.org/article/view/v059i10>

Karl W. Broman & Kara H. Woo, *Data Organization in Spreadsheets*, Vol. 72, Issue 1, 2018, The American Statistician. <https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989>

Workshop Materials – <https://github.com/RebeccaOlson/UCL_Workshops>

Our goal is to create a machine readable spreadsheet.

**Structuring data in spreadsheets**

The cardinal rules of using spreadsheet programs for data:

1. Spreadsheets are meant to be machine readable
2. Put all your (**variables in columns)** - the thing you're measuring, like 'weight' or 'temperature'.
3. Put each (**observation) in its own row**.
4. One piece of (information) per cell. **Don't combine multiple pieces of information in one cell**. Sometimes it just seems like one thing, but think if that's the only way you'll want to be able to use or sort that data.
5. **Leave the raw data (raw)** - don't mess with it! Make a copy of the spreadsheet and work in the copy.
6. Export the cleaned data to a (**text based) format** like CSV. This ensures that anyone can use the data, and is the format required by most data repositories.
7. Also remember that a spreadsheet must also be machine readable.

**Blank Cells vs. Zero value – are they the same?**

Open the **IDH\_Data** file.

How many sections do you see?

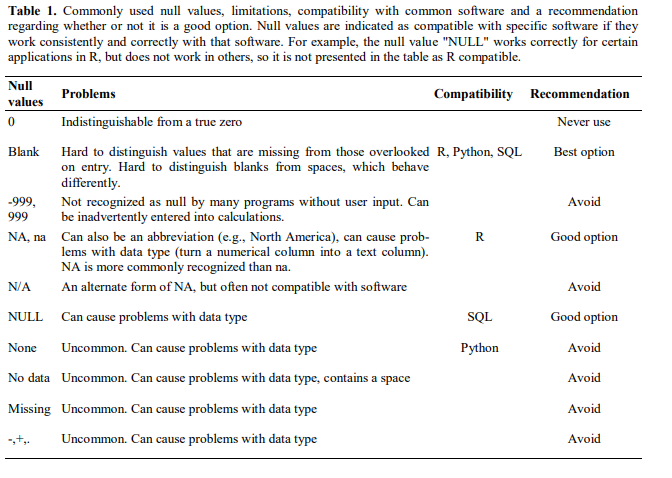
What formatting problems do you see?

For Excel we can use the term NULL to indicate that a value for the variable was not taken. Excel will also process a blank as a Null value. Other programs treat this expression and blank cells differently and you should be cautious when working between programs. The table below indicates what expression is best for each program.

Null Values Table From White et al, 2013, [Nine simple ways to make it easier to (re)use your data.](https://ojs.library.queensu.ca/index.php/IEE/article/view/4608)

Ideas in Ecology and Evolution - doi: 10.4033/iee.2013.6b.6f

Null Values.



**Metadata**

The information about the data – context, full information.

|  |
| --- |
| Open **IDH\_Data**  What metadata do you need to understand the variables? |

**Dates:**

**Note:** Dates are especially problematic for Excel. Here is a good article about the issues with date standards and how Excel causes issues.

Briney, Kristin A.. 2018. "The Problem with Dates: Applying ISO 8601 to Research Data Management." Journal of eScience Librarianship 7(2): e1147. <https://doi.org/10.7191/jeslib.2018.1147>

Download/open the **POD\_DATA\_for\_EPP** file.

Note the dates.

What do you notice?

Best practice tip = separate out the date into month, day, years.

|  |
| --- |
| Step 1. Highlight the Admission Date column. Check the format (should be DATE – select the correct formatting).  Step 2. Insert 3 empty columns next to the “Admission Date” column. On the top, enter Month, Day, Year  Step 3. Make sure all three columns are formatted as numbers, with 0 decimal places.  Enter in the Month column in the first data cell, =**MONTH(A2).** Look for the little box in the lower right corner of that cell. Drag it down to apply the formula to the whole column. Repeat with **=DAY(A2), =YEAR(A2)**  Try entering just 11/1.  What do you notice? |

**Data Validation:**

For more information, see Microsoft’s tips for applying validation to cells. <https://support.office.com/en-us/article/Apply-data-validation-to-cells-29FECBCC-D1B9-42C1-9D76-EFF3CE5F7249>

**QA for Spreadsheets**.

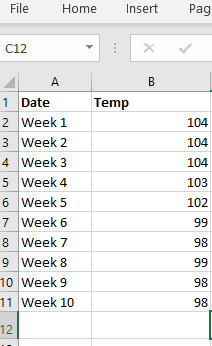
You can require a range or list for entries to a cell through using the Data Tools settings on Excel.

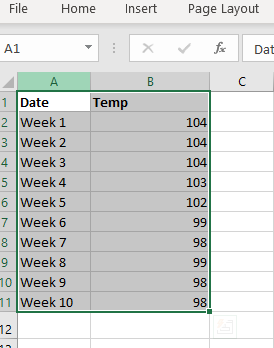
Best practice tip: set up range/list parameters BEFORE data entry.

|  |
| --- |
| **POD\_Data\_for\_EPB\_workshop** spreadsheet  **For numbers**:  Select the column H – **POD 1 # of times Up with Assistance Completed**.  From the **DATA** tab, select **DATA TOOLS**, **DATA VALIDATION**/**VALIDATION TOOLS**.  Under **Settings**, select “Whole number” from the **ALLOW** menu.  Select Minimum (0) and Maximum (3).  Under **Input Message** enter “Invalid Number” as the title, and an error message.  Click OK.  Try entering a value outside of 1-30 in the column.  **For entries on a list**:  Select the column **POD 1 Up with assistance Ordered**  From the **DATA** tab, select **DATA TOOLS**, **DATA VALIDATION**/**VALIDATION TOOLS**.  Under **Settings**, select “List” from the **ALLOW** menu.  In the **Source** field, enter the values you want to be accepted, separated by commas.  Enter **Yes, No**  Create an **INPUT** message then click ok. |

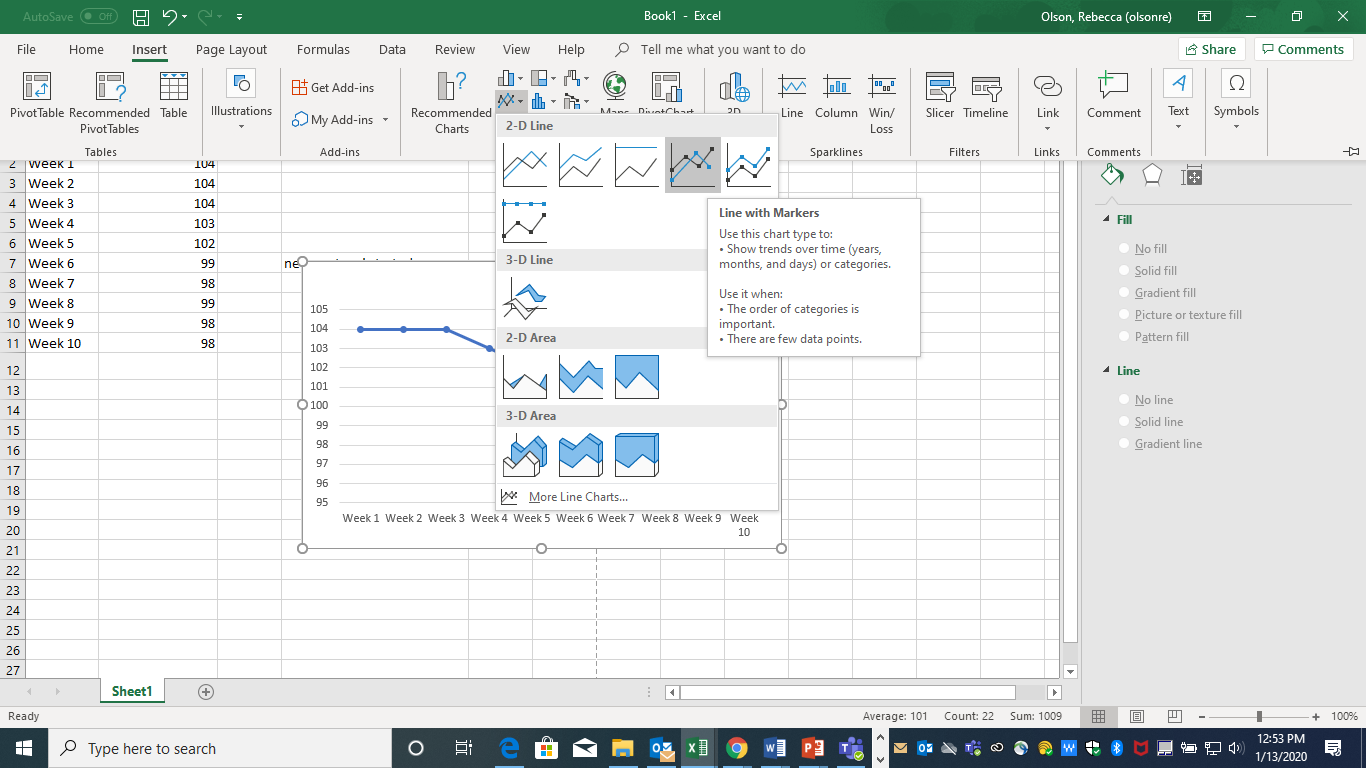
# Building a Run Chart Evidence Based Practice Workshop

**Step 1.** Open Excel. Enter your data into Excel Columns. Column A should be your time periods, Column B should be the data you collected. Give the two columns names and enter the data.

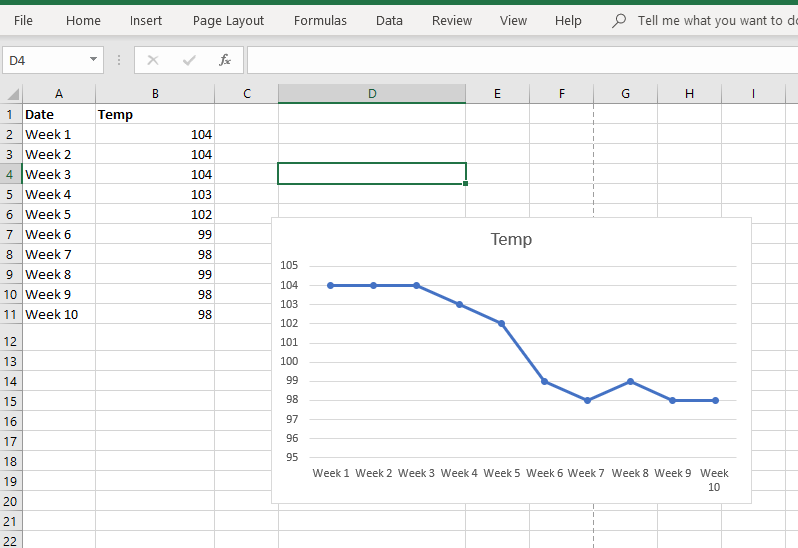


**Step 2.** Highlight all of the data by using your mouse

**Step 3**. From the INSERT menu at the top, select 2-D line with Markers.

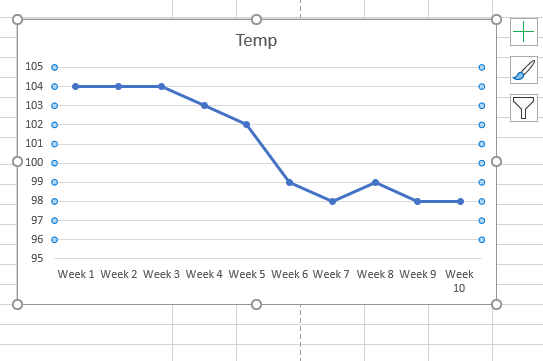


**Step 4**. You now have a basic Run Chart that you can format.

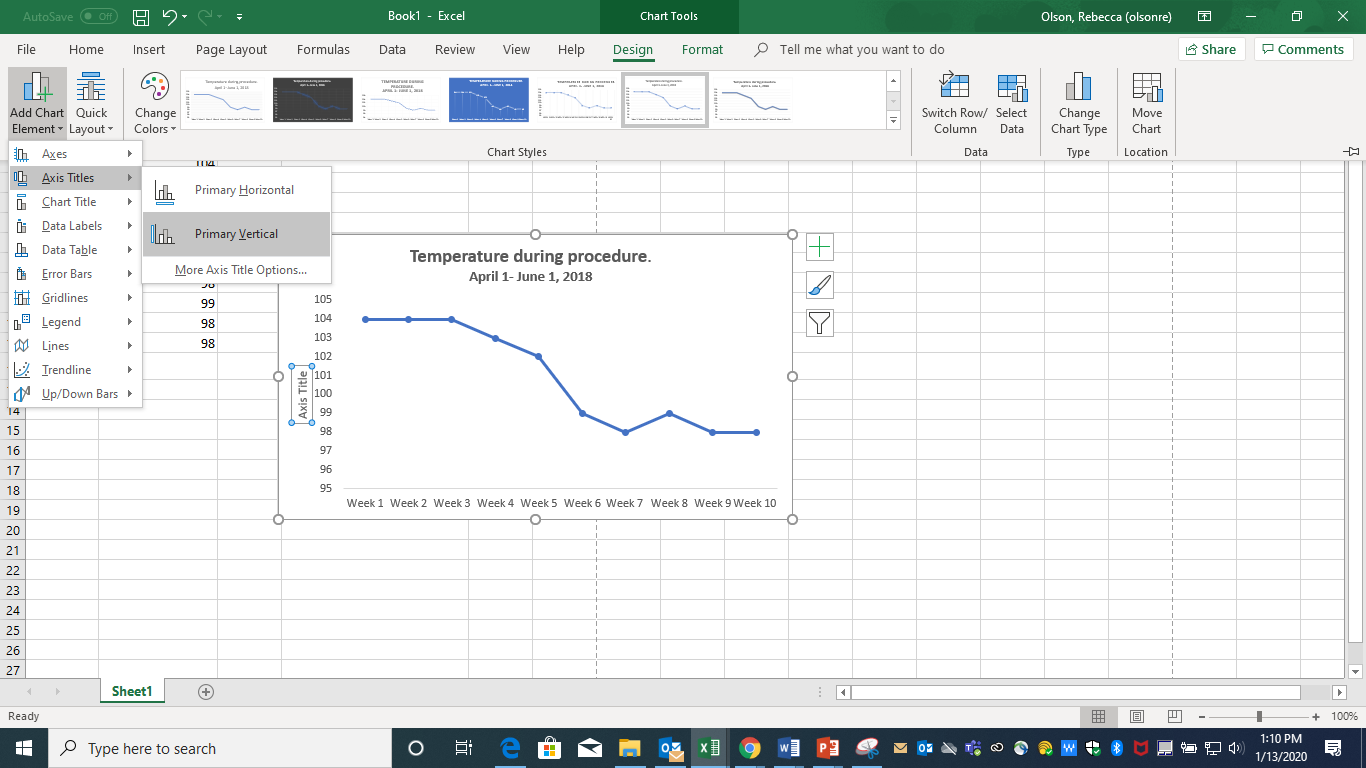


**Step 5.** Formatting depends upon what you want to do with your chart. Each version of Excel may have slightly different directions for formatting.

1. **Remove grid lines**. Since you may want to include a median line, you will want to remove the grid lines. Click on one of the grid lines and then hit “delete.”

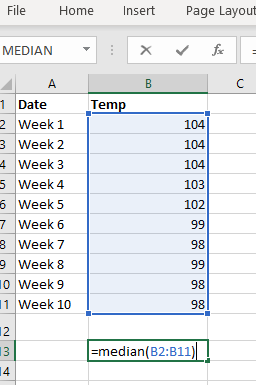


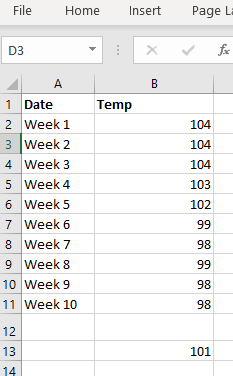
1. **Format the Vertical axis.** In this case, for temperature, you probably don’t need to change the range. But a good rule of thumb is to extend the range above and below the current range to include future changes. Some advise the data to be in the middle of the chart, some advise there to be 1/3 above/below of white space.
2. **Give your chart and your vertical axis a title**. Click where the title is and start typing. You can use formatting, add dates, etc. Make sure your vertical axis makes sense. Use the Design menu and add chart element for the Y and X axis.



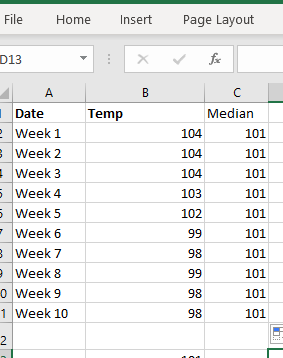
**Step 6.** **Add a median line**. In order for run charts to be analyzed, they require a median line. Most of the time, the line will be the median of all points being plotted. The median is the number in the middle if all the plot points were arranged from smallest to largest. You need to calculate the median of the plots.

Click on a blank cell in the spreadsheet (B13 in the image below). Type in **=median(B2:B11)** (or whatever the range of your points are). Click Enter. The median is 101.

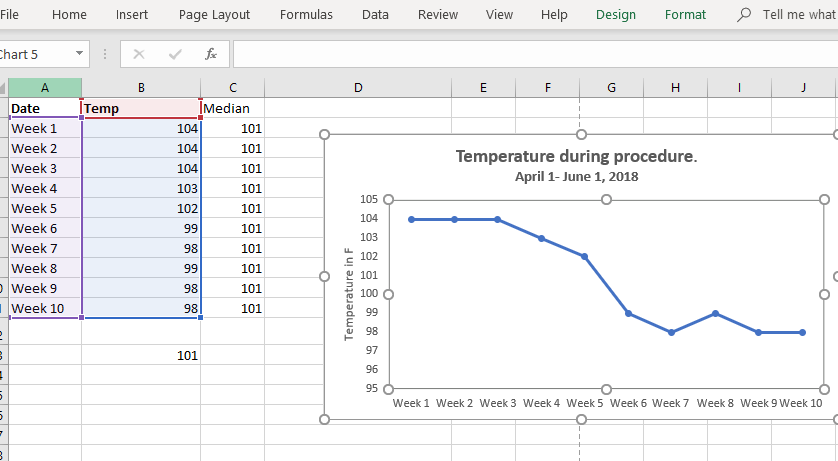


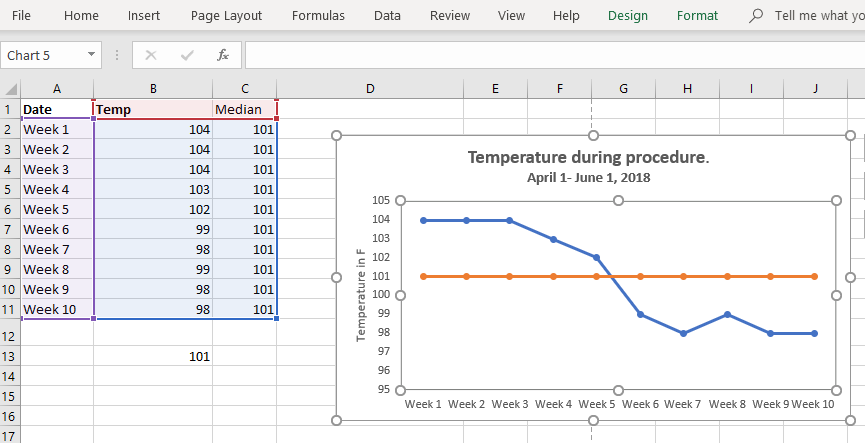


**Step 7**. Adding the median line to the chart. Name Column C **Median** and enter the value. Place the value in the entire column by dragging the **+**.



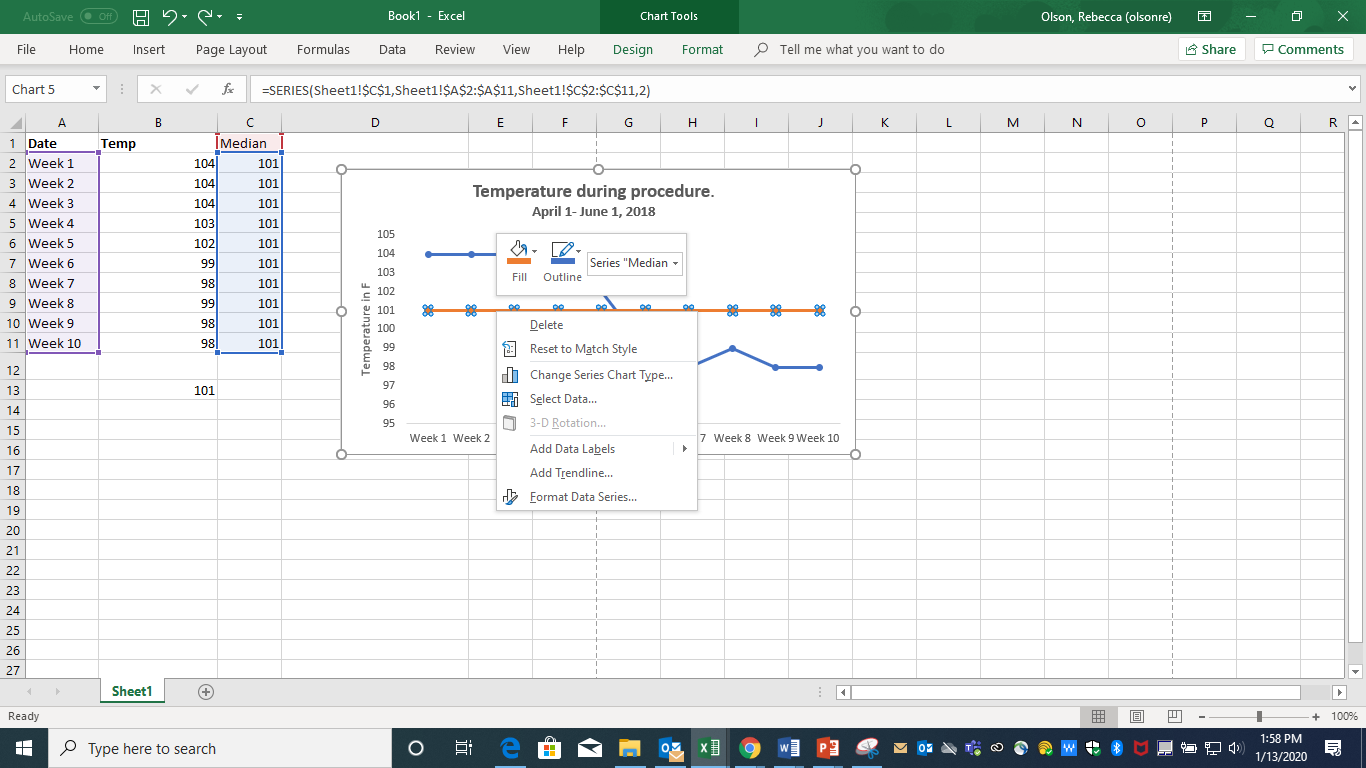
**Step 8.** Add the Median line to the run chart. Click on the graph. Note that Column A and B are highlighted. Drag the bottom right corner over column C. A new line will show up on the graph.



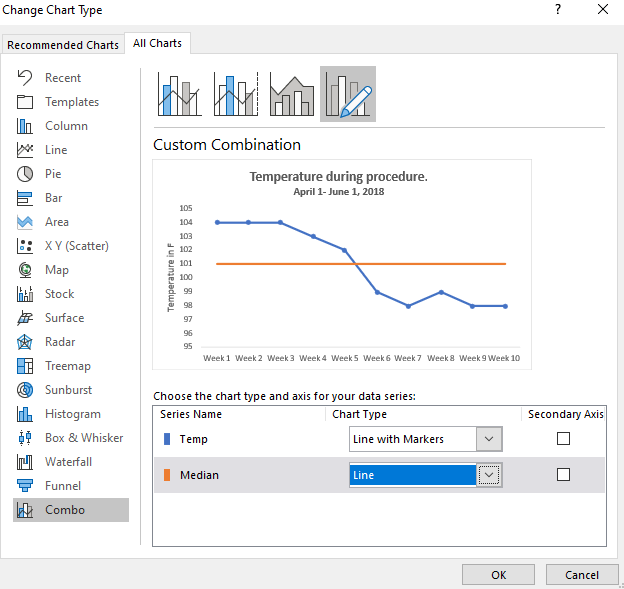
New Chart Line 

The median line should be smooth – so you will need to remove the formatted tick marks.

**Right click on the Median Line**. Chose “change chart style” and select the “plain line.”



**Select Line (no markers) for the median.**



**7 steps in constructing a Run chart (adapted from Provost & Murray, 2011)**

1. State the question that will be answered. Start collecting the data
2. Horizontal Scale
   1. Usually a time scale in normal increments (hours, days, weeks, months, quarters, etc)
   2. Best practice tip: Label Future increments, even if you don’t have the data yet. Scale should cover future graph requirements, not just current data
3. Vertical Scale
   1. Make the vertical scale easy to plot, read, and remember to leave space for future data that may have significant range changes.
   2. Criteria for a good scale:
      1. Most data is located in the middle half of the graph
      2. Labeled values on the axis should be round numbers, equally spaced.
      3. Unlabeled tic marks should be easily read and plotted
      4. Ratio should be 2.5 Vertical to Horizontal
      5. Estimate range of data points to be plotted on vertical scale and develop vertical scale with high and low points enough for variation in data.
4. Plot
   1. Plot data points with a dot or another symbol
   2. You can connect the points with a line, but DATA is the DOTS, not the line
5. Label the graph completely with a useful title
   1. Horizontal scale – sequence of data (month, case, etc)
   2. Vertical scale – name of measure or characteristic being studied
6. Calculate the Median of the data
   1. Median (middle number of values) should be calculated from the data on the run chart and be placed on it.
   2. Can change median/add additional median if values plotted change significantly.
7. Additional Info
   1. Can add a goal or target line if appropriate
   2. Can annotate important events (treatment 1 started, change 2 on this date)

**Using your data**

Open up **IDH\_Data**

What sorts of information could you chart?

What was the percentage of IDH symptoms presented?

You need to find the COUNT of the entries and then find the percentage

At the end of column B, Nausea Y or N, **=COUNTA(B3:B52)**

Will get you the total count of answers.

Insert a new tab and name it Documented Symptoms

Copy the top columns for this section to that sheet.

Copy the counts found and then **past special – values**.

Format the cells as % (with 0 decimals). Change the numbers into percentages **=(B2/50)**

Insert a chart.

**Exercise on your own:**

Try making charts using the **Documented Interventions** and **MD Ordered Interventions**. Put each in a new tab.

Will you be doing counts? Averages? Percentages?

Advanced exercise – compare either of the cleaned Post data spreadsheets with the Pre data spreadsheets and figure out how you might make a run chart.

What issues/limitations do you find?

Run Chart Resources used.

Anhøj J, Olesen AV (2014) Run Charts Revisited: A Simulation Study of Run Chart Rules for Detection of Non-Random Variation in Health Care Processes. PLOS ONE 9(11): e113825. <https://doi.org/10.1371/journal.pone.0113825>

Canadian Foundation for Healthcare Improvement (2014) How to Make a Run Chart in Excel

<https://www.cfhi-fcass.ca/sf-docs/default-source/on-call/22-10-14-how-to-make-a-run-chart-in-excel-cfhi-tutorials-en.pdf?sfvrsn=3ffdd44_2>

Oregon Association of Hospitals and Health Systems (2018)How to Make a Run Chart Using Excel

<https://www.oahhs.org/assets/documents/files/how-to-make-a-run-chart-using.pdf>

Provost, L. P., & Murray, S. (2011). *The health care data guide : Learning from data for improvement*. Retrieved from [https://ebookcentral.proquest.com](https://ebookcentral.proquest.com/) See especially, Chapter 3 on Run Charts

<http://uclid.uc.edu/record=b5782001~S39>

|  |
| --- |
| **Additional help:**  Often googling your issue will turn up exact answers either in video form or on a web page.  Linked-In Learning has several tutorials on Excel. UCIT provides access to it for free.  <https://www.uc.edu/about/ucit/services/linkedin-learning-access.html>  UC’s library catalog – search in Books for Excel – many, many online books with tutorials.  Contact us and we will work with you. |

Contact us at [ASKDATA@UC.EDU](mailto:ASKDATA@UC.EDU)

Research & Data Services,

University of Cincinnati Libraries

<https://libraries.uc.edu/rds>