# NM2207

Session 08 Codealong

# Overview of what we will do today:

* Data
* Charts

**Part 1**

Revision of concepts

**Spreadsheet tricks (Chapter 2 of Hands-on Data Viz)**

* Filtering data
* Subsetting data
* Coloring cells

**Aim of this practice:** In this segment, we are going to practice how to do the following: **Filtering data**, **subsetting data** and **coloring cells**. The dataset we will be using for this activity has been uploaded into Canvas.

* Filtering data
  + There are only 2 variables in this dataset.
    - Volume of mobile data use
    - Time (in terms of quarter years)
  + Let’s filter the data so that we are only looking at mobile data use for the first and last quarter of the year (such that we can compare if the volume of mobile data use has increased or decreased between the start and end of the year – ignore 2004 and 2019 as these data are incomplete).
    - First, we select column A, then click on data, and then filter. A tiny triangle will now appear beside the cell. From here we can select the data we want to keep.

Graphical user interface, application, table, Excel

Description automatically generated

* Subsetting data
  + Now, let’s say we want to be even more specific with the data we are interested in. Let us only look at the years where the volume of mobile data consumed (in petabytes) exceeds 1 petabyte per quarter.
  + Using the filter tool, how would you do this?
    - While we can manually select and de-select the values individually as we did previously, the fastest way would be to make use of excel’s abilities. When prompted with what you would like to filter, select the option greater than 1 to keep all values greater 1 in your spreadsheet, while filtering out the values that are less than 1.

Graphical user interface, application

Description automatically generated

* Coloring cells
  + Now, we are left with only the years where mobile data consumption exceeds 1 petabyte per quarter year. However, it is still not very easy to visualize the differences in mobile data consumption between each time period. As such, we will be using another amazing excel function, “color scales”.
    - Select Column B, then we click on Home, and find the conditional formatting tool.
    - Select color scales and pick the first option.
    - Now, your data can be easily visualized, and we see a upward trend in mobile data usage both within years and across years (the closer the color is to red, the lower the value is. The closer the color is to green, the greater the value is).

**Graphical user interface, application, table, Excel

Description automatically generated**

* Visualize this in a different way in Excel. Select the rows and columns. Insert->Trendline. Is this what you want to show in your web page?

**Part 2**

**Summary of learnings (Chapter 2 of Hands-on Data Viz)**

* Finding a dataset
* Manually preparing it for exploration
* Filtering, subsetting, color-coding
* Charting
* Reading a dataset from the internet (gapminder.org)
  + Go to <https://www.gapminder.org/data/>
  + Click on “choose individual indicators”
  + After you choose one, “download as csv” on the right
* Cleaning the dataset (focus on a few countries and two columns)
  + Filter the data:
    - Choose only 8 countries, Singapore, Malaysia, China, Philippines, Indonesia, Thailand, Vietnam, USA
  + Subset the data:
    - Choose only 11 years, 2001 to 2011.
  + Using colors to see trends
    - Select the column
    - Conditional formatting: “scale”
  + We want to compare how fast the countries have progressed in the last ten years. Use functions to calculate things
    - Calculate the increase in 10 years between first and last column
* Update the happinessbarChart.html with your new numbers!

**Reflection**

* Based on the chart you drew in Part 2, what have you learned about these eight countries? Does this surprise you? Why/why not?
* What are the errors you encountered today, and how did you address them? Add a table showing this in the html body of your page.