

Homework 2

Use the PubMed dataset (you can use library) and download at least five years of scientific articles.

1. For each of the following terms: *Influenza*, *Obesity*, *Cancer*, *Covid-19* create an area chart and compare their frequencies for each of the following years: 2016, 2017, 2018, 2019, 2020. For example, The keyword frequency of Influenza: 2016- 134, 2017-32, 2018-120,...
The keyword frequency of Cancer: 2016:230, 2017:122, ...
Then, frequency of these four keywords should be plotted together in one area chart.

2. Please use a dumbbell-chart and report the changes in following keywords for 2020 and 2019: *Influenza*, *Covid-19*, *Depression*, *Mental health*, *Physical activity*, *Wearable*

3. Get Covid-19 statistics from the city of Massachusetts:
<https://www.mass.gov/doc/covid-19-cases-in-massachusetts-as-of-april-13-2020/download>
or any other city, country you like.

Then, design a Choropleth map and visualize (number of confirmed cases or percentage of positive tests) on the map of Boston or MA. It is ok, if you manually enter the data into your code.

You are free to choose any other state or city, it is not limited to Boston or MA.

You need to prepare a report on your tasks and findings along with a video file describing what you have done. You can copy paste your codes, its results and your description into a Word document, Python Notebook or you can use R notebook.

Your deadline for delivering this home work is written on the blackboard online. Please feel free to ask your question and prepare it for presentation for the next session.

Hint: Good link to learn using map:
<http://bcb.dfci.harvard.edu/~aedin/courses/R/CDC/maps.html>