PREREQUISITES YOU SHOULD COMPLETE

MSc PUBLIC HEALTH DATA SCIENCE
This document is a compendium of resources about prerequisite concepts you should be familiar with before beginning the Master program in Public Health Data Science. We encourage you to read and/or train with the materials listed below.

**BIOSTATISTICS**

- **General statistical theory concepts**
  
  Read through the 6 chapters of this interactive online book: [http://students.brown.edu/seeing-theory/](http://students.brown.edu/seeing-theory/)

- **Introductory statistics & basics of R programming**
  
  Following [http://swirlstats.com/students.html](http://swirlstats.com/students.html) instructions, install Rstudio, R and the swirl package, and complete the following courses in the given order:
  1. R Programming
  2. Exploratory Data Analysis
  3. Getting and Cleaning Data
  4. Statistical Inference
  5. Regression Models
  6. Statistical Inference

- **Deepen R capabilities**
  
  « R for data-science » book (free online version): [https://r4ds.had.co.nz/](https://r4ds.had.co.nz/)
  

  An additional resource on *statistical hypothesis testing* can be found on the *Statistics in Action with R* website: [http://sia.webpopix.org/statisticalTests1.html](http://sia.webpopix.org/statisticalTests1.html). Read through it, then play around with the Shiny application at [http://shiny.webpopix.org/sia/testMean/](http://shiny.webpopix.org/sia/testMean/).

- **Basic ideas about causality**
  

- **Deepen statistical prerequisites**
  
  - Follow the « Statistical Inference » class from Johns Hopkins University on Coursera: [https://www.coursera.org/learn/statistical-inference](https://www.coursera.org/learn/statistical-inference). The online book from Lauren Cappiello « Introduction to statistics » [https://bookdown.org/lgpcappiello/IntroStats/random-variables.html](https://bookdown.org/lgpcappiello/IntroStats/random-variables.html) is also a very useful resource blending key statistical theoretical concepts with R without too much complexity and a lot of explanations.
EPIDEMIOLOGY

- Basic concepts
  - Epidemiological concepts compulsory in **basic tools**:
    - Introduction the most common types of epidemiological study designs (experimental, cross-sectional, cohort, case-control).
    - Introduction to bias (selection, information, confounding) and effect modification.
    - Differences between measures of health status (prevalence, incidence, risk) and association (risk ratio, rate ratio, odds ratio).

- Fundamentals of Epidemiology
  - **ActivEpi Web** is a multimedia electronic free book, created by David Kleinbaum (doctor and professor of epidemiology) that provides an interactive ressource to learn the fundamentals of epidemiology, biostatistics, and an introduction to mathematical modeling.
    - Log In access: [https://activepi.herokuapp.com/](https://activepi.herokuapp.com/)

- Recommended readings
  - For students with intermediate background in epidemiology: **Szklo M, Nieto FJ. Epidemiology: beyond the basics. Burlington: Jones & Bartlett Publishers; 2014.**

MEDICAL INFORMATICS

- Notions of relational database design and implementation
  - Normalisation
  - First, Second and Third Normal Forms
  - Entity-relationship modelling
  - Relational modelling
  - SQL language

You can follow this course: [https://www.coursera.org/learn/database-management](https://www.coursera.org/learn/database-management) or this one: [https://www.udemy.com/relational-database-design/](https://www.udemy.com/relational-database-design/)
• **Object-oriented programming**

Data structure and algorithm  
Object-oriented design: classes and objects, inheritance, polymorphism, encapsulation  
Basics of Python programming: basic instructions, introduction to Object Oriented programming with Python.

Enroll in the following udemy course (until OOP in Python section):  
[https://www.udemy.com/python-masterclass-for-beginners/](https://www.udemy.com/python-masterclass-for-beginners/)  
It could be completed by the following course on OpenClassRoom  