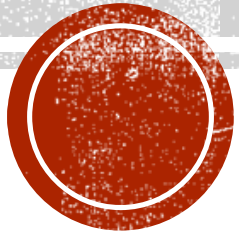
DATA 606 CAPSTONE IN DATA SCIENCE

A VIEW ON HCV DATA

SRIHARSHA DAPARTI

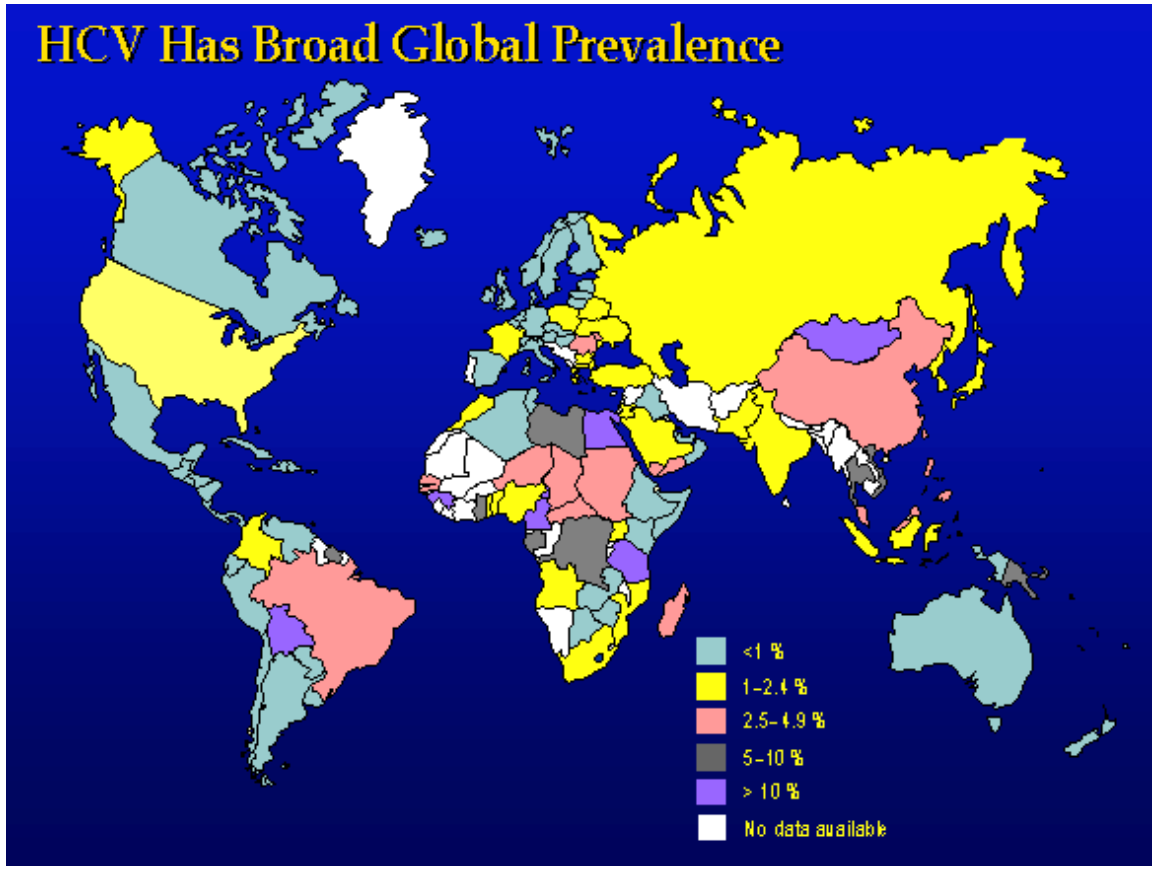
UX46295

Instructor : DR. ERGUN SIMSEK



INTRODUCTION

- Hepatitis C virus (HCV) is a major health problem worldwide. In 2015, the global prevalence of HCV infection was 1.0%, with the highest prevalence in the Eastern Mediterranean Region (2.3%) followed by the European one (1.5%). The annual mortality due to HCV-related complications is estimated to be approximately 700000 deaths.



ABOUT THE DATASET

- **Hepatitis C Virus (HCV) for Egyptian patients**. This data was obtained from UCI Machine Learning Repository
  Citation: Dua, D. and Graff, C. (2019). Link.
- The dataset has two files. The first is the dataset itself which shows the anonymous records of Egyptian patients who underwent treatment dosages for HCV about 18 months and the second file contains the discretization parameters for each and every attribute in the first file.
- The dataset contains about 1000 patient records with 29 attributes for each record explaining the treatment .
- The attributes of the patient records are:

• Age	• WBC Count – White Blood Cell Count	• ALT 48 – ALT Week 48
• Gender	• RBC Count - Red Blood Cell Count	• ALT after 24 w – ALT after 24 weeks
• BMI	• HGB - Haemoglobin	• RNA Base
• Fever	• Platelets Count	• RNA 4
• Nausea/Vomiting	• AST 1 – Aspartate Transaminase ratio	• RNA 12
• Headache	• ALT 1 – Alanine Transaminase ratio Week 1	• RNA EOT – RNA at End Of Treatment
• Diarrhea	• ALT 4 – ALT Week 4	• RNA EF – RNA Elongation Factor
• Fatigue & Generalized bone ache	• ALT 12 – ALT Week 12	• Baseline Histological Grading
• Jaundice	• ALT 24 – ALT Week 24	• Baseline Histological Staging
• Epigastric Pain	• ALT 36 – ALT Week 36	

INTENDED WORK

To see **covariance** between symptoms with results of tests done at different weeks and determine the **prominent features** that contribute to extreme histological staging i.e., Cirrhosis.

To try and answer all the questions that the dataset might pose during different levels of analysis

The dataset presents a **classification problem.** So the goal is to come up with a model that can **predict the stage of disease** using few or all features.

REFERENCES

- Omran D, Alboraie M, Zayed RA, et al. Towards hepatitis C virus elimination: Egyptian experience, achievements and limitations. *World J Gastroenterol*. 2018;24(38):4330–4340. doi:10.3748/wjg.v24.i38.4330
- World Health Organization (WHO) 2016. Hepatitis C Fact sheet. July.
- Ministry of Health and Population/Egypt, El-Zanaty and Associates/Egypt, and ICF International. 2015. Egypt Health Issues Survey 2015. Cairo, Egypt: Ministry of Health and Population/EgyptandICFInternational.