**INTERNATIONAL UNIVERSITY**

**VIETNAM NATIONAL UNIVERSITY, HCM CITY**

**School of Computer Science & Engineering**

****

**GROUP PROJECT REPORT**

**DATA SCIENCE & DATA VISUALIZATION**

**Course: Data Science & Data Visualization**

**Lecturer: Tran Thanh Tung**

**Topic: An overview of COVID 19 in the world**

**Group members:**

* Bui Xuan Hieu ITDSIU18047
* Phan Thi Nhu Yen ITDSIU18048
* Nguyen Viet Anh ITDSIU18027

**TABLE OF CONTENTS**

CHAPTER-1: INTRODUCTION

1.1 Introduction to the covid19 topic……….………………...4

1.2 Link to github repositions……………….……………….4

1.3 Background and motivations……….……………………5

* + 1. Background…………………………………………5
    2. Motivations…………………………………………5

1.4 Objective of the project……………………………………5

* + 1. Primary objective
    2. Secondary objective

CHAPTER-2: ANALYSIS AND DESIGN

2.1 Data and data processing…………………………………6

2.2 Visualization designs……………………………………1

2.3 Coding implementation…………………………………2

2.3.1 Coding

2.3.2 Project screenshots

CHAPTER-3: CONCLUSION AND LESSON LEARNT

* 1. Project limitation…………………………………………9
  2. Conclusion………………………………………………10
  3. Lesson learnt……………………………………………12

3.4 References………………………………………………10

.

*(This page is intentionally left blank)*

**CHAPTER-1: INTRODUCTION**

* 1. **Introduction to the corona virus – what is it?**

A group of people standing in front of a crowd

Description automatically generated

Figure 1. People protect themselves and the others by wearing a mask in public places in corona virus pandemic.

“Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.

Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment.  Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.” (WHO, 2020)

**Globally**, as of **10:40am CEST, 16 June 2020**, there have been **7,924,527 confirmed cases** of COVID-19, including **434,367 deaths,** reported to WHO.

* 1. **Link to github repositions**

<https://github.com/weeblord2030/DSDV-PJ>

**1.3 Background and motivations**

**1.3.1 Background**

As of learning Data Science and Data Visualization course at International University - VNU, through all the theory and lab sessions, these taught us all about what it means to create a deceptive visualization and more importantly, how to create an effective one.

**1.3.2 Motivations**

The main motivations are:

* To fulfill our understandings for Data Science and Data Visualization course through efficiently visualizing covid data.
* To apply fundamental principles and techniques for our project, and though the covid data is complex, we have learnt how to handle the basics.
* This corona pandemic is the new normal state that our world is facing, although Vietnam has gone to the “quarantine phase” and the pandemic is now controlled. In many countries in the world, this is still a hot topic and people are definitely very concerned about it.
  1. **Objective of the project**

**1.3.1 Primary objectives**

* To design visual data analysis solutions for corona virus pandemic - our world on-going problem.
* The primary questions to answer are:
* In each chart we have different questions answered: that surround mostly the data of the world
  + 1. **Secondary objectives**
* To design a web-based interactive visualization
* To evaluate and critique visualization designs
* To implement interactive data visuaizations (web development skills) on the group project
* Improve our D3.js skills, along with a little HTML, CSS and Java script.

**CHAPTER-2: ANALYSIS AND DESIGN**

**2.1 Data and data processing**

There are many awesome sites and trackers out there already. Johns Hopkins University has an awesome [COVID-19 Dashboard](https://coronavirus.jhu.edu/map.html) that compiles data from a variety of sources and displays just about any statistic you could be interested in regarding the global pandemic.

**Data Source 1:**COVID-19 json and csv data from JHU CSSE repository. [http://coronavirusdashboard.live](http://coronavirusdashboard.live/)

**A screenshot of a social media post

Description automatically generated**

We decided to focus on data about the number of global confirmed cases. We used Johns Hopkins University open GitHub repository of confirmed global by day in the world. This data came in the form of a CSV file called time\_series\_covid19\_confirmed\_global.csv.

**Data Source 2:** COVID-19 json and csv data from WHO website. <https://www.who.int/data/collections>

This data came in the form of a CSV file called Data3%2C4%2C5.csv and world\_case 31-05.csv.

**Data Source 3:** world json and csv data from the internet and from labs materials

Lastly, we used world’s data to visualize demographic trends called

**2.2 Visualization designs**

**2.3 Coding implementation and project screenshots**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

A screenshot of a map

Description automatically generated

**CHAPTER-3: CONCLUSION AND LESSON LEARNT**

**3.1 Project limitation**

We are still work on how to efficiently:

- Having a more interactive and efficient story telling visualizations.

- More good-looking and informative charts and collect them together in a web page.

* 1. **Conclusion**

When we are given the freedom to pick topic and our works in this project is basically “we get out what we put in” type of project. We rely mainly on our lab sessions, therefore the visualizations are still not quite come across together in solely one web-based. We will put it in the future enhancements.

**3.3 Lessons learnt**

## Incremental progress: Making small improvements each day was, by far, the most satisfying part of this project. Since we started from square one, we really had to take it one feature at a time.

The open source community so powerful to be able to play with demos, clone repositories, and ask for help in online forums so that we, as a community, can push forward into new domains without ever having to reinvent the wheel. Huge shoutout to the anonymous heroes on Stack Overflow who got us through this project.

We have the chance to improve our D3.js skills, along with a little HTML, CSS and Java script through implementing interactive data visualizations (web development skills) on the group project.

**3.4 References**

# 1. COVID-19 Movement & Cases Tracker with Demographic Trends

(2020, May 20nd). Retrieved from <https://sevans09.github.io/covid-movement-vis/>

2. Coronavirus pandemic (covid-19) – the data. Retrieved from <https://ourworldindata.org/coronavirus-data>

3. COVID-19 json and csv data from WHO website. <https://www.who.int/data/collections>