



Understanding Missing Data and How to Deal with it

Haw-minn Lu, PhD
Principal Data Scientist, West Health Institute
Lecturer, University of San Diego

21th Python in Science Conference (SciPy 2022)
July 12, 2022
Austin TX

Getting Started

Open this document at <https://github.com/WestHealth/scipy2022-missingness-tutorial>. Scan this QR Code



Installation

- Docker
 - Secure
 - `docker pull westhealth/scipy2022-missingness-tutorial:latest`
 - `docker run -p8888:8888 westhealth/scipy2022-missingness-tutorial:latest`
 - If you wish a different port change the first 8888
 - goto <http://localhost:8888> and paste hash from your standard out
 - Insecure
 - `docker pull westhealth/ scipy2022-missingness-tutorial:insecure`
 - `docker run -p8888:8888 westhealth/scipy2022-missingness-tutorial:insecure`
 - goto <http://localhost:8888> no hash is required

Installation

- Colab
 - Get Notebooks and data
 - Clone from Git
 - `git clone https://github.com/WestHealth/scipy2022-missingness-tutorial.git`
 - Download Zip
 - [https://github.com/ WestHealth/scipy2022-missingness-tutorial](https://github.com/WestHealth/scipy2022-missingness-tutorial)
 - CODE -> Download ZIP
 - Upload notebooks in colab directory to <https://colab.research.google.com/>
 - Create missingness_tutorial folder in your google drive
 - Upload helpers.py and data/ folder to that directory

Thank You

- Resources
 - References provided at the end of each notebook
 - This workshop will be morphed into a blog series at Westhealth's Data Science Blog
 - <https://westhealth.github.io/>
- Questions
 - Post at issue at <https://github.com/WestHealth/scipy2022-missingness-tutorial/issues>
 - This could help others who have a similar question
 - Email me:
 - hlu@westhealth.org
 - hlu@san Diego.edu