

Week-3 Prep

September 4, 2020

1 Week-3: Data Visualization and Cleaning

1.1 Readings:

- From Think Like a Data Scientist:
 - Chapter-4: Data Wrangling: from capture to domestication
 - Chapter-5: Data Assessment: Poking and Prodding

Some questions to keep in mind as you are reading:

- What is Data Wrangling?
- Why wrangling isn't a task with can be described exactly?
- Think about a sport and choose two very important figures of this sport (Messi vs Maradona for soccer, M. Jordan vs L. James in basketball, Sampras vs Federer for tennis, etc.) and ask yourself how would you make a good (good in the context of data science) comparison of these two players.
- Where you would go to get data?
- What might be some challenges of the data you collected?
- What are the some types of messy data?
- What might go wrong with the data when you collect it from an online source?
- What are the common file formats?
- What are the common pitfalls in data wrangling?
- What are the some of the most important descriptive statistics?
- Compare descriptive and inferential statistics.
- What is histogram?
- When we talk about the distribution of a variable what do we mean?

1.2 For Coding Part:

- [Python Data Science Handbook](#)
 - Chapter-3: Data Manipulation with Pandas
 - Chapter-4: Visualization with Matplotlib

1.3 Additional Resources

- [matplotlib Gallery](#)
- [PyData Conference - Cleaning and Tidying Data in Pandas](#)
- [Data School - Top 25 Pandas Tricks](#)
- [Become a Python Analyst - Pandas Tutorial](#)