# W1

## Intro of DS

* What is data science?
  + Data science is finding the pattern and law behind the data, get inside information from data
* What compare with traditional algorithms how ML different?
  + Traditional al are based on formula, but ML is based on phenomenon, and the phenomenon is based on multiple
* List four V of big data
  + Velocity, Varity, volume, veracity (volume is size of data, veracity is uncertainty of dataset)
* What is the size of small data? Size of big data?
  + Size is over how many GB, 10 GB is big data
  + Why we focusing on small first?
    - Because the big data questions form by small dataset questions
* Why use R (2)?
  + R focus on data science and it has interactive environment
* What is rectangular data?
  + Table with column name/kind, multiple row with
  + What is non rectangular data?
    - Videos, text, graph
* What Is Hypothesis?
  + Hypothesis generation?
    - Brain storm ideas, questions about
  + Hypothesis confirmation?
    - Using hypothesis
* Why we focus on hypothesis generation?
  + How it helps in comprehending business problem?(4 reasons)

## Data Science life cycle

* List 6 stages and question for each stage
  + For each stage why we need them?

## Intro of R

* Different between arrow (<-) and equal?
  + Answer: (<-)use for apply value, = use in function parameter passing
* What is naming restriction?(4)
* Definition of?
  + Vectors
  + Matrices
  + Arrays (what is different between vector and array?)
  + List( different between it and vector?)
  + Data Frames