

HUDK 4050:CORE METHODS IN EDM



<https://stackoverflow.com/help/how-to-ask>

Research

- Search for your answer

Research

- Search again for your answer

Research

- Search again for your answer just to make sure

Clarity

- Ask question clearly
- Ask only one one question at a time
- Make sure the title ***says what it is on the tin***
 - “R confusion” **(Bad)**
 - “What is the difference between a factor variable and a numeric variable in R?” **(Better)**
- Make sure spelling and grammar are correct (ask someone if you are unsure)
- Identify code with by using `` and >
- Don't use tabs! (SO doesn't interpret them)

Format

- Title
- Question body: expand on title
 - How did the problem arise, brief background
- Code
- Reiterate question
- Thank you

Reproducible Example

<https://stackoverflow.com/help/mcve>

- **Minimal:** Use as little code/explanation as possible
 - Don't post your whole assignment!
 - Recreate the problem with a smaller toy data set (you may solve the problem by doing this)

Reproducible Example

<https://stackoverflow.com/help/mcve>

- **Complete:** Include all aspects of problem
 - Where did the data come from?
 - A single problem
 - What is the overall goal and the specific goal of the code
 - If you have a lot of code you need to isolate the line with the issue
 - It was working and then stopped

Reproducible Example

<https://stackoverflow.com/help/mcve>

- **Verifiable:** Can the problem be reproduced
 - Toy data
 - Code
 - Any other relevant information: system, R version

Asking Good Questions

- Life skill not just for SO
- Many people don't have it
- It takes practice
- Is worth spending time to think about it

K-means Gotchas

- Assumes there are clusters to find - it will find clusters regardless of whether there are any or not
- Does not work on some shapes (Like PB&J need an even spread)
- Need uniform scale (uniform variance) - larger scale will swamp smaller scale
- Doesn't work on categorical data of more than two categories (and the scale may be difficult to interpret)
- Can get stuck on local minima (need to run iterations)
- Too easy