Lab 2

Everything Data

CompSci 216 Spring 2019
Seating chart

Find your group on Sakai
1. Site Info
2. Groups you are a member of
3. Lab-XX

Back of the room

Front of Room
Read-Eval-Print Loop (REPL)

• Data science requires human-in-the-loop interactions with various datasets
  – Come up with a command/query to execute
  – Run it
  – See the result
  – If it isn’t exactly what you want, go back to step 1

• Systems like Postgres and Python provide a “Read-Eval-Print Loop” (REPL) to support this process
Jupyter

• [http://jupyter.org/](http://jupyter.org/)

• Jupyter is
  – A more interactive, user-friendly way to REPL
  – A way to share your methodology and results with others
  – A way to turn in your homework

• The basic unit of work in Excel is a “spreadsheet”; in Jupyter, it’s a “notebook”
Pull Up Today’s Lab

cd ~/CS216-s2019-READONLY

git pull

• You will see a folder CS216-s2019-READONLY/labs/lab02

• The assignment is in jupyter notebook lab02.ipynb
How to Run Jupyter

• NOTE: assumes that you have run everything needed to do Homework 2 already on your VM

• Completely within VM
  – Log into VM via Remote Desktop Connection
  – Open LXTerminal and run: `jupyter notebook`
  – A URL will appear – copy and paste it into Chromium

• Using web Browser on local machine + Jupyter on VM
  – Connect to VM via SSH tunneling:
    ```
    ssh -L 8888:127.0.0.1:8888 vcm@vcm-XXXX.vm.duke.edu
    ```
  – Inside the SSH session, run: `jupyter notebook`
  – A URL will appear – copy and paste it into Chromium
How to Run Jupyter

• NOTE: assumes that you have run everything needed to do Homework 2 already on your VM

• Completely within VM
  – Log into VM via Remote Desktop Connection
  – Open LXTerminal and run: jupyter notebook
  – A URL will appear – copy and paste it into Chromium

• Using web Browser on local machine + Jupyter on VM
  – Connect to VM via SSH tunneling:
    `ssh -L 8888:127.0.0.1:888 vcm@vcm-XXXX.vm.duke.edu`
  – Inside the SSH session, run: jupyter notebook
  – A URL will appear – copy and paste it into Chromium
Jupyter Cells

• Each cell has a “formula” in a “language”
  – Code: in this case, Python code
  – Markdown: text meant for human consumption

• A cell can be “evaluated”
  – In [X] (formula being evaluated)
  – Out [X] (result of evaluating that formula)
  – Until “X” in “Out[X]” is a number, evaluation has not yet finished
  – “X” indicates the order cells were evaluated in
    • If you run cells in a non-linear order, the X’s will be out of order

• Evaluate (or re-evaluate) a cell by clicking on the cell and pressing Ctrl+Enter
• Clear the entire notebook’s execution via Kernel > Restart & Clear Output