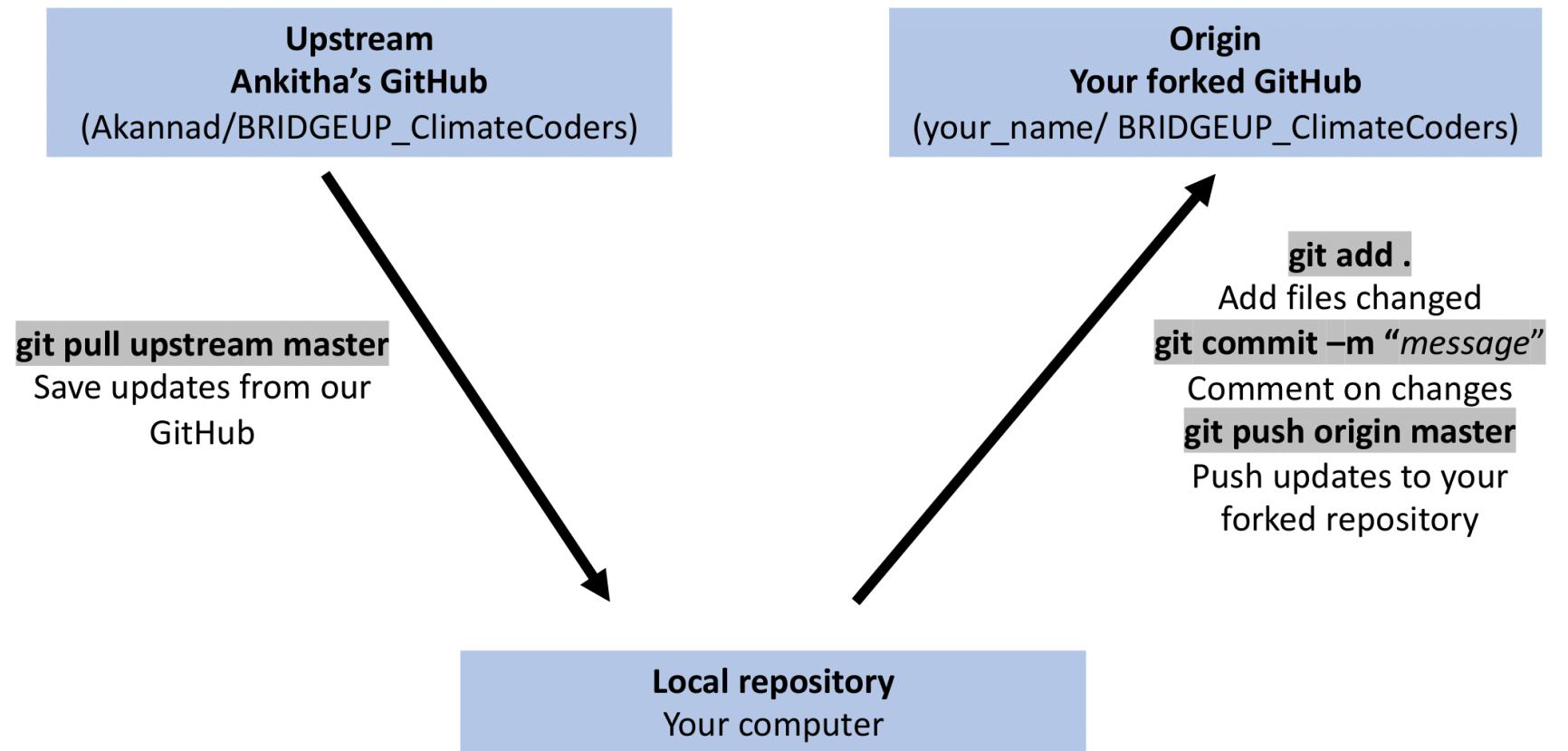


HOUSEKEEPING

- Zoom guidelines:
 - Mute unless you need to speak
 - Share errors on Slack
 - Short description of error and what you tried out already
 - Error message
 - Section of code/ terminal



PANDAS REVIEW

Sort values from maximum to minimum:

- `pandas_dataframe.sort_values(by, ascending)`
 - **by = column_name** name of column to sort
 - **ascending = True** values will be in ascending order
 - **ascending = False** values will be in descending order

Find all values that meet some condition:

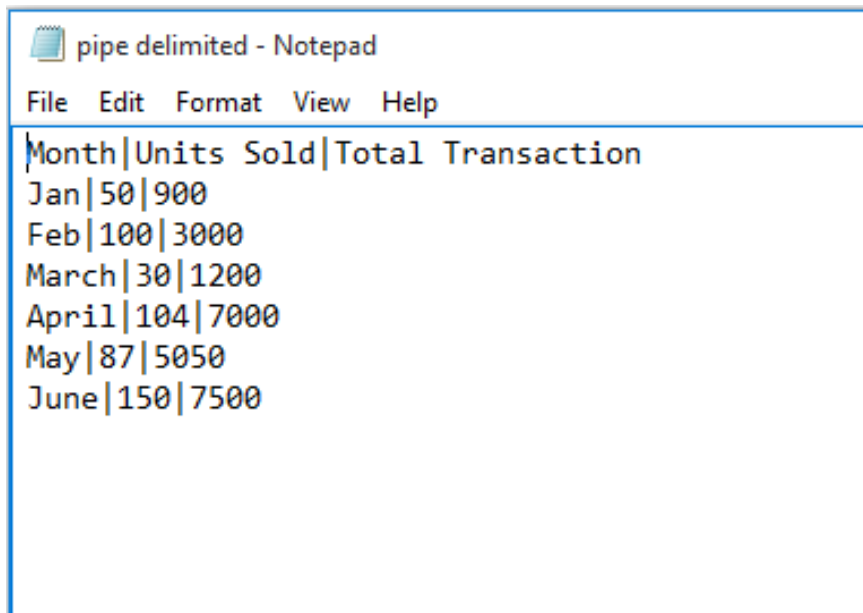
- `pandas_dataframe.where(cond)`
 - **cond** save all values that meet the condition and replaces others with NaN

PROCESSING DATA

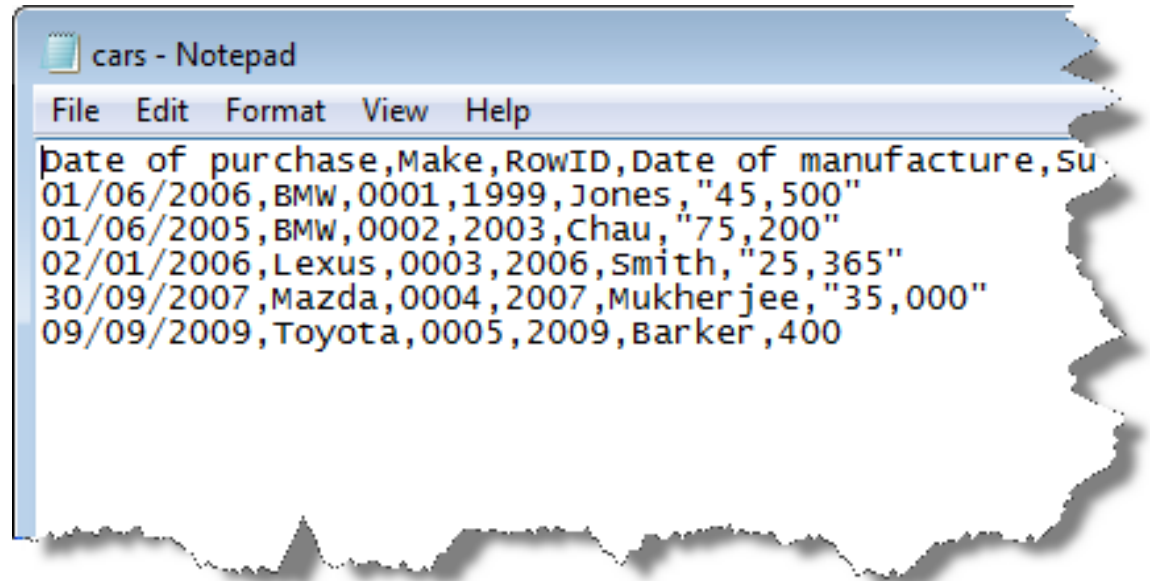
- **Make it into format we can read into Python**
- Make into a format we can interpret
- Fill in any missing gaps (if possible)
- Slice - only look at certain time periods, regions, etc.

DELIMITED TEXT FILES

- Data in a .txt file that is separated by “something” called the delimiter
- Comma, colon, space etc = delimiter
- CSV or **comma separated values** → type of delimited text file



```
pipe delimited - Notepad
File Edit Format View Help
Month|Units Sold|Total Transaction
Jan|50|900
Feb|100|3000
March|30|1200
April|104|7000
May|87|5050
June|150|7500
```



```
cars - Notepad
File Edit Format View Help
Date of purchase,Make,RowID,Date of manufacture,Su
01/06/2006,BMW,0001,1999,Jones,"45,500"
01/06/2005,BMW,0002,2003,Chau,"75,200"
02/01/2006,Lexus,0003,2006,Smith,"25,365"
30/09/2007,Mazda,0004,2007,Mukherjee,"35,000"
09/09/2009,Toyota,0005,2009,Barker,400
```

PLAN TO PROCESS DATA

Is there any missing data?

Do we need to convert our data to a format we can easily understand?

What functions and libraries might we need?