Big Data lab

Class exercise 3

Browse the data

```
str()      # structure
head()     # 6 first observations in the variable
class()    # the class of the variable
dim()      # the dimensions of the variable
summary()  # summarize (context dependent)
table()    # create a contingency table
xtabs()    # create a contingency table
```

install packages into R

```
install.packages('package-name')
```

Plot to file

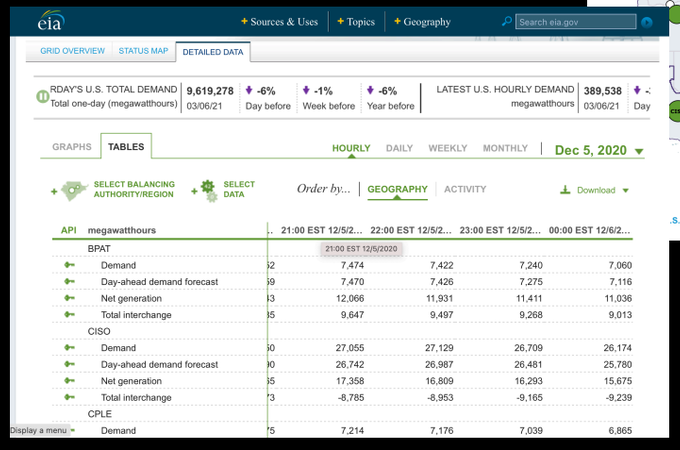
```
pdf('file-name.pdf')      # open a file for writing
plot(…)                   # graphics commands
plot(…)
lines(…)
hist(…)
…
dev.off()                 # finalize and close the file
```

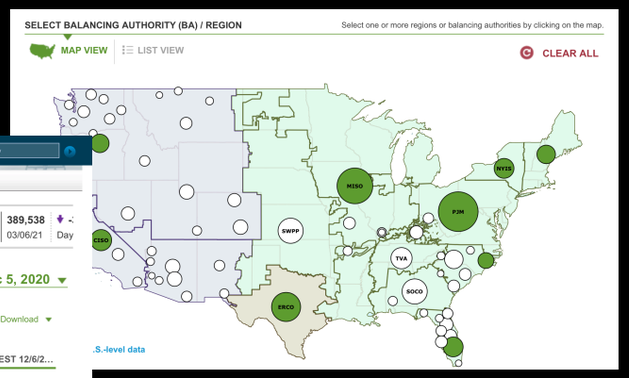
Organize your data

- Directory structure
- Data away from code
- git lfs

Business Measurements

Power consumption in USA





https://www.eia.gov/realtime_grid/#/data/table?end=20201205T00&start=20201128T00&bas=003g0j02000g5

OLAP Assignment

Create Business measurements

- User experience with multidimensional linear regression analysis…
- In the code
  - Load the data
  - Rearrange the data
  - Estimate means, variance
  - Calculate regression fit
- Your task
  - Create a data cube
  - Estimate the linear fit in a slice across time and location

OLAP Assignment

Q: What is the mean daily power generation across the US?

- Consider **the week of 7 Feb. 2021** (7-14 Feb).
- Output - A chart (X-axis = day, Y-axis Net generation). Draw a line to give the mean value

- Q: What is the minute power demand in the east coast?
  - Locations **PJM**, **NYIS**, **ISNE**, **FPL**, **CPLE**
  - Time **10:00-18:00**
  - Time **20:00-03:00**
  - Output - A chart (X-axis = Time, Y-axis = Demand)

Please submit working code **Week3_power.r** and a **pdf** file **Week3_power.pdf** with two plots on two pages