

# 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

**eia** Sources & Uses Topics Geography Search eia.gov

GRID OVERVIEW STATUS MAP DETAILED DATA

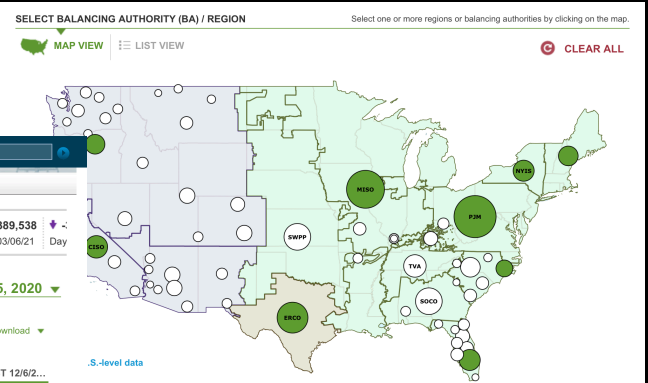
RDAY'S U.S. TOTAL DEMAND 9,619,278 -6% -1% -6% LATEST U.S. HOURLY DEMAND 389,538 - Day  
Total one-day (megawatthours) 03/06/21 Day before Week before Year before megawatthours 03/06/21 Day

GRAPHS TABLES HOURLY DAILY WEEKLY MONTHLY Dec 5, 2020

+ SELECT BALANCING AUTHORITY/REGION + SELECT DATA Order by... GEOGRAPHY ACTIVITY Download

API	megawatthours	21:00 EST 12/5/2...	22:00 EST 12/5/2...	23:00 EST 12/5/2...	00:00 EST 12/6/2...
BPAT					
Demand	32	7,474	7,422	7,240	7,060
Day-ahead demand forecast	39	7,470	7,426	7,275	7,116
Net generation	13	12,066	11,931	11,411	11,036
Total interchange	35	9,647	9,497	9,268	9,013
CISO					
Demand	30	27,055	27,129	26,709	26,174
Day-ahead demand forecast	30	26,742	26,987	26,481	25,780
Net generation	35	17,358	16,809	16,293	15,675
Total interchange	13	-8,785	-8,953	-9,165	-9,239
CPLE					
Demand	15	7,214	7,176	7,039	6,865

Display a menu



[https://www.eia.gov/  
realtime\\_grid/#/data/table?  
end=20201205T00&start=202011  
28T00&bas=003g0j02000g5](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