Financial Programming and Databases

WANG, Zigan (王自干)
Ph.D., Columbia University
Faculty of Business Economics, The University of Hong Kong

What you need to know about this course

- It’s very important if you are interested in data
  - Work with machine/data
  - Work with text/writing
  - Work with people/social relationship

- The style of this course
  - In-class practices
  - Friendly atmosphere
  - Win-win outcome

Who should and shouldn’t take this course

- Interested in data processing work in the future

- Who shouldn’t take this course:
  - Already very proficient in web-scraping (selenium), multiprocessing and Pandas
  - Already very familiar in CAR, Tobit or Logit, Gephi, Compustat, CRSP etc.
    - You are overqualified
  - Take too many courses and don’t have much time for projects and practice
  - Likely to complain about the workload

```
import pandas as pd
import numpy as np
import os
from datetime import date
from copy import copy
from nameparser import HumanName as hn
from functools import partial

DATA_PATH = os.pardir+os.sep+os.pardir+os.sep+'data'+os.sep+'director_data'+os.sep
FILE_PATH = os.pardir+os.sep+os.pardir+os.sep+'files_needed'+os.sep
year_list = range(1992, 2016)

def today(string):
    return date.today().strftime('%Y%m%d_')+string

#CUSIP_Year_left_joining_director_number_TRI_change_SP1500
#no data preloading
"""
append near tri
input:  data/director_data/CUSIP_Year_left_joining_director_number_diff_years_SP1500.csv
output: data/director_data/CUSIP_Year_left_joining_director_number_TRI_change_SP1500.csv
"""
left_joining = pd.read_csv(DATA_PATH+'20160405_CUSIP_Year_left_joining_director_number_diff_years_SP1500.csv', index_col=0)
levels = [1,2,5,10]
tri_path = os.pardir+os.sep+'data'+os.sep+'plant_opening_closing'+os.sep

tri_change = pd.Series(map(lambda x: pd.read_csv(tri_path+'{:s}num_TRI_change_near_CUSIP_{:d}_mile.csv'.format('20160403_', x), dtype={'cusip': str}, index_col=0), levels), levels)

for i, mile in tri_change.iteritems():
    print i
    mile.index = mile.index.map(lambda x: x.zfill(9))
    left_joining['TRI_change_{:d}_mile'.format(i)] = left_joining[left_joining['year']!=2015].apply(lambda cy: mile.ix[cy['cusip'], str(cy['year'])], axis=1)

left_joining.to_csv(DATA_PATH+today('CUSIP_Year_left_joining_director_number_TRI_change_SP1500.csv'), index_col=0)
```

TABLE 2—SORTING BETWEEN RESIDENTS AND PROGRAMS

	log NIH fund (MD) (1)	Median MCAT (MD) (2)	MD degree (3)	DO degree (4)
log NIH fund (major)	0.3724***	0.0154***	0.0462***	0.0025
	(0.0119)	(0.0007)	(0.0032)	(0.0022)
log NIH fund (minor)	0.1498***	0.0084***	0.0208***	0.0048*
	(0.0137)	(0.0008)	(0.0040)	(0.0028)
log number beds	−0.0972***	−0.0021	−0.0104	−0.0098**
	(0.0221)	(0.0014)	(0.0064)	(0.0045)
Rural program	−0.0687	−0.0040	−0.0010	0.0138*
	(0.0437)	(0.0027)	(0.0117)	(0.0082)
log case-mix index	0.1894**	0.0136**	0.4670***	0.0574***
	(0.0940)	(0.0058)	(0.0255)	(0.0179)
log first-year salary	0.0126	0.0590***	0.3001***	0.0969***
	(0.1717)	(0.0106)	(0.0467)	(0.0327)
log rent	0.4612***	0.0727***	0.1811***	−0.0012
	(0.0600)	(0.0037)	(0.0168)	(0.0118)
Observations	10,842	10,872	23,984	23,984
R^2	0.1318	0.1282	0.0381	0.0079

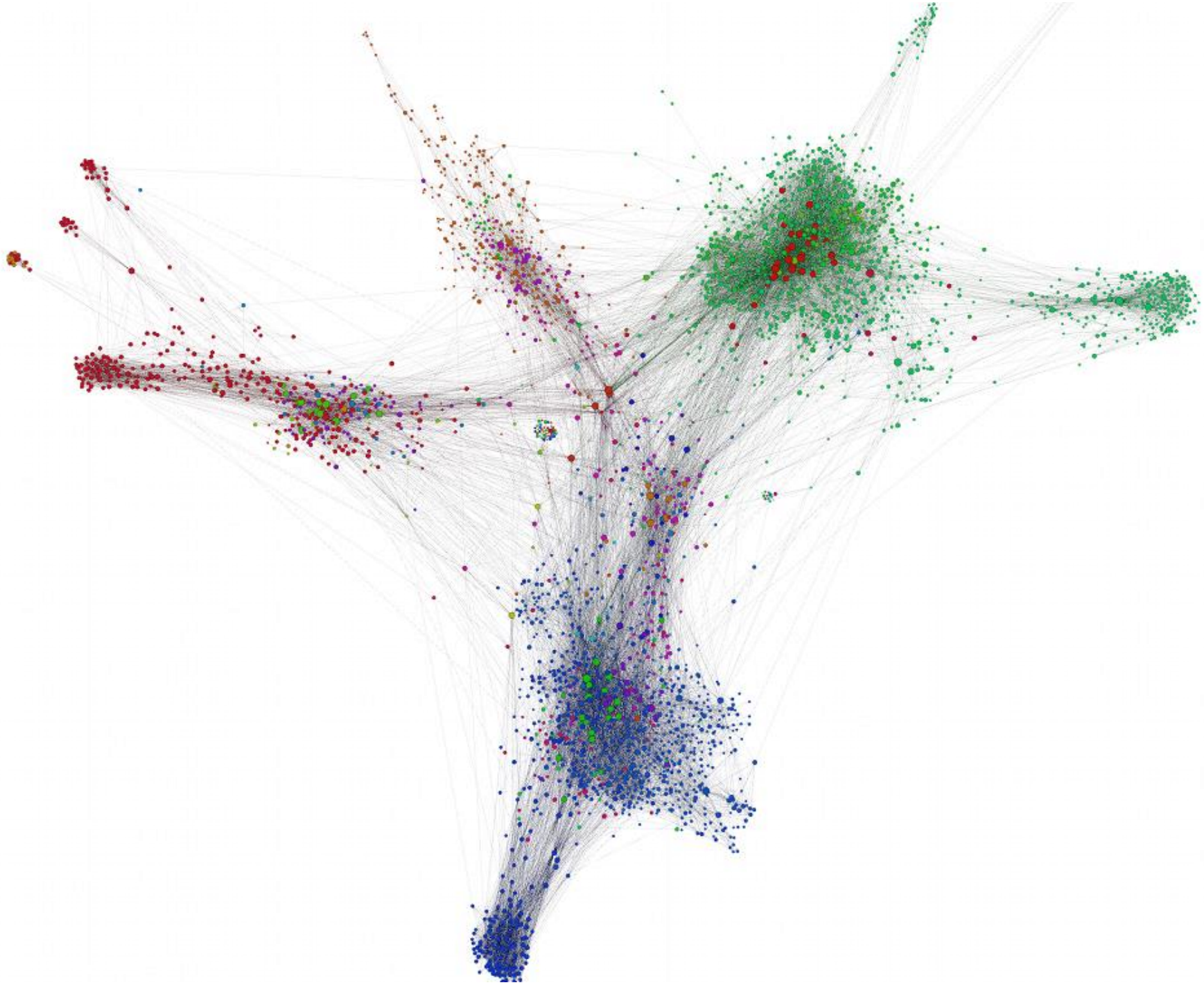
Notes: Linear regression of resident characteristic on matched program characteristics. Column 1 restricts to the residents from medical schools with positive NIH funding. Column 2 restricts to the residents institutions reporting MCAT scores in the medical school admission requirements in 2010–2011. All specifications include indicators for programs with no NIH funding at major affiliates, no NIH funding at minor affiliates, and a missing Medicare ID for the primary institution. Standard errors in parentheses.

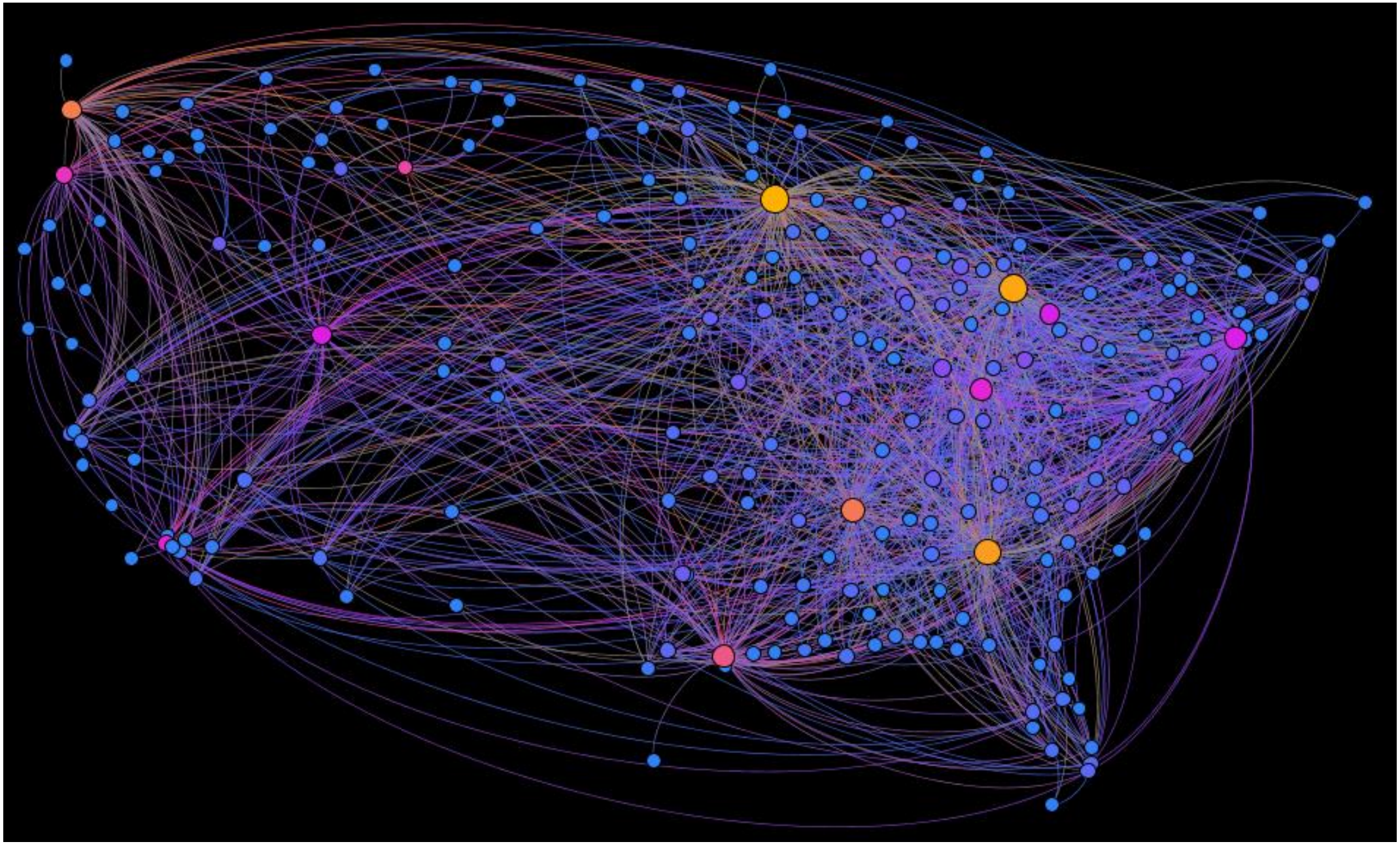
***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Status	StatusCode	Date_Annou~d	Year_Annou~d	Date_Effec~e	Acq_Name	Acq_NameClean	Acq_Primar~C	Acq_State	Acq_State_~r	AcqCUSIP	Acq_Ticker
Withdrawn	2	9/8/1981	1981		Southeast Banking Corp,Miami	Southeast Banking Corp	6021	Florida	FL	841338	STB
Completed	1	5/3/1982	1982	6/28/1983	InterFirst Corp,Dallas,Texas	InterFirst Corp	6021	Texas	TX	458916	
Completed	1	6/2/1983	1983	1/1/1984	First Florida Banks Inc	First Florida Banks Inc	6021	Florida	FL	320264	FFBK
Withdrawn	2	7/26/1983	1983		First City Bancorp of Texas	First City Bancorp of Texas	6712	Texas	TX	319594	FBT
Completed	1	9/1/1983	1983	9/20/1985	Hartford National Corp, CT	Hartford National Corp	6021	Connecticut	CT	416655	HNAT
Withdrawn	2	11/3/1983	1983		Mellon National Corp	Mellon National Corp	6021	Pennsylvania	PA	585518	
Completed	1	2/13/1984	1984	12/18/1985	Bank of New England,Boston,MA	Bank of New England	6021	Massachusetts	MA	63840	BKNE
Withdrawn	2	4/5/1984	1984		Affiliated Bankshares of CO	Affiliated Bankshares of CO	6022	Colorado	CO	8182	AFBK
Withdrawn	2	4/9/1984	1984		United Virginia Bankshares Inc	United Virginia Bankshares Inc	6022	Virginia	VA	913164	UVBK
Withdrawn	2	4/9/1984	1984		First Virginia Banks Inc,VA	First Virginia Banks Inc	6022	Virginia	VA	337477	FVB
Completed	1	4/23/1984	1984	2/1/1985	Norstar Bancorp,Albany,NY	Norstar Bancorp	6021	New York	NY	656538	NOR
Completed	1	4/23/1984	1984	2/8/1985	Bancorp Hawaii Inc,Honolulu,HI	Bancorp Hawaii Inc	6022	Hawaii	HI	59685	BNHI
Completed	1	5/21/1984	1984	12/27/1984	Zions Utah Bancorp,UT	Zions Utah Bancorp	6022	Utah	UT	989722	ZION
Completed	1	7/2/1984	1984	7/1/1985	Sun Banks Inc,Orlando,Florida	Sun Banks Inc	6022	Florida	FL	866635	SU
Completed	1	9/4/1984	1984	4/9/1985	Bank of Virginia Co	Bank of Virginia Co	6022	Virginia	VA	65446	BKV
Completed	1	9/11/1984	1984	3/14/1985	Marine Corp, Milwaukee,WI	Marine Corp	6022	Wisconsin	WI	568236	MCRP
Completed	1	10/16/1984	1984	7/1/1985	First Commerce,New Orleans,LA	First Commerce	6021	Louisiana	LA	319779	FCOM
Completed	1	10/18/1984	1984	8/20/1985	First Commerce,New Orleans,LA	First Commerce	6021	Louisiana	LA	319779	FCOM
Completed	1	10/25/1984	1984	7/26/1985	Key Banks Inc	Key Banks Inc	6021	New York	NY	493067	KEY
Withdrawn	2	11/28/1984	1984		United Jersey Bks,Princeton,NJ	United Jersey Bks	6021	New Jersey	NJ	910748	UJB
Completed	1	12/5/1984	1984	5/1/1985	Midlantic Banks Inc,Edison,NJ	Midlantic Banks Inc	6022	New Jersey	NJ	597806	MIDL
Withdrawn	2	12/5/1984	1984		First National State Bancorp	First National State Bancorp	6021	New Jersey	NJ	335742	FNS
Completed	1	12/12/1984	1984	10/8/1985	First Wisconsin Corp	First Wisconsin Corp	6021	Wisconsin	WI	337570	FWB
Status Unknown	3	1/9/1985	1985		First National State Bancorp	First National State Bancorp	6021	New Jersey	NJ	335742	FNS
Withdrawn	2	1/18/1985	1985		Comerica Inc	Comerica Inc	6022	Michigan	MI	200340	CMCA
Completed	1	1/22/1985	1985	8/22/1985	Barnett Banks FL Inc,Florida	Barnett Banks FL Inc	6022	Florida	FL	68055	BBI
Status Unknown	3	1/23/1985	1985		Louisiana Bancshares Inc	Louisiana Bancshares Inc	6022	Louisiana	LA	546192	LABS
Completed	1	1/25/1985	1985	1/25/1985	Boatmen's Bancshares,St Louis	Boatmen's Bancshares	6022	Missouri	MO	96650	BOAT
Completed	1	2/20/1985	1985	8/30/1985	Citizens and Southern GA Corp	Citizens and Southern GA Corp	6021	Georgia	GA	173124	CSGA
Status Unknown	3	2/25/1985	1985		CBT Corp,Hartford,CT	CBT Corp	6021	Connecticut	CT	124850	CBCT
Completed	1	2/26/1985	1985	8/6/1985	Security Bancorp,Southgate,MI	Security Bancorp	6022	Michigan	MI	813771	SECB
Completed	1	3/4/1985	1985	11/29/1985	First Union Corp,Charlotte,NC	First Union Corp	6021	North Carolina	NC	337358	FTU
Completed	1	3/20/1985	1985	3/20/1985	SouthTrust Corp,Birmingham,AL	SouthTrust Corp	6021	Alabama	AL	844730	SOTR
Completed	1	3/29/1985	1985	12/30/1985	United Virginia Bankshares Inc	United Virginia Bankshares Inc	6022	Virginia	VA	913164	UVBK
Status Unknown	3	4/3/1985	1985		Equimark Corp,Pittsburgh,PA	Equimark Corp	6022	Pennsylvania	PA	294432	EQK
Withdrawn	2	4/19/1985	1985		BankAmerica Corp	BankAmerica Corp	6021	California	CA	66050	BAC
Completed	1	5/8/1985	1985	12/16/1985	Indiana National Corp	Indiana National Corp	6712	Indiana	IN	455011	INAT
Status Unknown	3	5/9/1985	1985		First Wisconsin Corp	First Wisconsin Corp	6021	Wisconsin	WI	337570	FWB
Status Unknown	3	5/9/1985	1985		First Wisconsin Corp	First Wisconsin Corp	6021	Wisconsin	WI	337570	FWB
Status Unknown	3	5/13/1985	1985		Citizens Fidelity, Louisville	Citizens Fidelity	6022	Kentucky	KY	174566	CEDY





Textbook that you don’t actually need

- **Title:** Python data analytics : data analysis and science using Pandas, Matplotlib, and the Python programming language

  Online access
- **Author:** Fabio Nelli author.
- **Author:** Shubham Singh Tomar; editor.
- **Subjects:** Python (Computer program language); Data mining; Database searching; Scripting languages (Computer science)
- **Description:** This book will help you tackle the world of data acquisition and analysis using the power of the Python language. Nelli provides coverage of pandas, an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. and shows the strength of the Python programming language when applied to processing, managing and retrieving information. This resource examines how to go about obtaining, processing, storing, managing and analyzing data using the Python programming language. Python and other open source tools can be used to wrangle data and tease out interesting and important trends in that data that will allow you to predict future patterns. Whether you are dealing with sales data, investment data (stocks, bonds, etc.), medical data, web page usage, or any other type of data set, Python can be used to interpret, analyze, and glean information. --
- **Related Titles:** Series: Expert's voice in Python.
- **Publisher:** New York : Apress
- **Format:** 1 online resource : illustrations.
- **Identifier:** ISBN9781484209585
- **Creation Date:** 2015
- **Language:** English

Python Data Analytics: Data Analysis and Science Using Pandas, Matplotlib, and the Python Programming Language

by Fabio Nelli

Apress © 2015 (364 pages) *Citation*
ISBN:9781484209592

By expertly showing the strength of the Python programming language when applied to processing, managing and retrieving information, this b

Recommend? yes no

Table of Contents

Rules

- You can leave the class but only during the break
- If you really need to leave the room, be quite
- Eating and drinking are allowed
- No attendance call
- No unprovoked complaint of workload or late drop-off, constructive suggestions are welcome
- No final exam, only projects
- Questions and discussions are welcome

- Again, we first need a friendly classroom

Grades

- 2017 fall semester, 21.28% of the whole class received A+, 8.51% received A, 70.21 received B+, no one below B+
- 2018 spring semester, over 50% in A range
- In general, most class participants performed very well - far beyond my expectation.
- Again, we first need a friendly classroom

Where does this style come from?

- Richard Clarida, vice chairman of the Federal Reserve

rhc2@columbia.edu

to zw2160

Here is the take home exam.

Rich

Where does this style come from?

- Joseph Stiglitz

- Patrick Bolton

This semester’s grade

- Depend on my overall evaluation of your performance
- I’m fair person. Even I didn’t like some people in previous classes, I gave them A+ because they deserved it.
- There is a benchmark for me to evaluate – workload points system for projects

Projects/Homework

Projects that I will assign

- All Python questions on Codeacademy.com
- Make a 10-minute presentation of Python code (examples and slides will be given)
- Write a web crawler. I will give 30 websites for you to choose but you can design your own.
- A data transformation problem set
- Data visualization of a social network using Gephi
- In-class test on Python programming on November 15

Course Outline

Course Outline

- Introduction of Python (2 weeks)
  - Anaconda
  - IDE – spyder
  - Print
  - Open and write
  - String, list, dictionary, tuple
  - If … then
  - For loop
    - While loop
  - Try … except (exception handling)
  - Function

Course Outline

- Data management with Python (3 weeks)
  - Packages
    - Numpy/Scipy
    - Dataframe
    - Pandas
  - OOP
    - Class
  - Geolocation
  - Multiprocessing
  - Pickle, Json, HDF5 (.h5)
  - Web scraping (regular expression, beautifulsoup, selenium)

Course Outline

- Database related software (1 week)
  - Cloud computing environment
  - Gephi

Course Outline

- Statistics (1 week)
  - OLS
  - Probit/Logit
  - Tobit
  - Poisson
  - Causality
  - IV

Course Outline

- Data (1 week)
  - Compustat
  - 6-, 8- and 9-digit CUSIP
  - CRSP
  - SDC
  - Others