

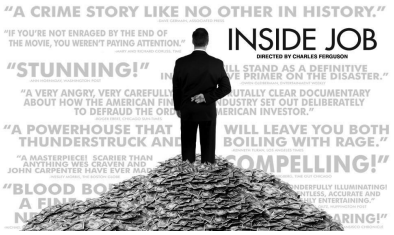

Stat 107: Introduction to Business and Financial Statistics
Class 1: Introduction





# We start with Matt Damon

- Harvard Alum Matt Damon




# Where Am I?

**Michael Parzen: Sci 300b**

**Office Phone: 617-495-8711**

**Email: mparzen@fas.harvard.edu**



# Who Am I?

HARVARD School of Public Health

CHICAGO BOOTH
The University of Chicago Booth School of Business

EMORY | GOIZUETA BUSINESS SCHOOL




# Course Website

- Course website: https://canvas.harvard.edu/courses/14476
- There you will find:
  - Syllabus
  - Administrative Announcements
  - Lecture Notes
  - R and Excel Tutorials
  - Assigned Homeworks
  - Other Study Material (web links, handouts, etc...)



# Lecture Notes

- Lecture notes will be handed out at the start of each class.
- They also will be available around 24 hours ahead of time on the course web site.

## Computing



- Computing will be a very important component of the class.
- We will be using Excel (a very small amount) and the free statistics package R.
- Details on downloading R are on the web site.
- We will not be using VBA in Excel (visual basic for applications). However, this is a useful skill to know.

## Grading



- 20% of your grade will be determined by homework assignments.
- 10% of your grade will be a final group project. More details about the project will be given towards the middle of the semester.
- 30% of your grade is based on a midterm (October 24, in class)
- 40% of your grade is based on a Final (December 13, 2pm)

## Collaboration Policy



- See web site for what is considered cool and uncool behavior.
- Based on CS50's code of conduct.

## Optional Weekly Sections



- Time and locations will be announced later this week as we learn how many students are in the course.
- You will not be sectioned; you can go to any section you want; there will even be a recorded weekly section.

## Guest Speaker



- We will have 1 guest speaker on Nov 28.
- Attendance is required, counts as a hw (excused Harvard absences are fine, of course).

## What this course is:



- Introduction to mathematical and statistical techniques used to solve investment problems.
- Learning how to program in R
- Basic concepts in modern investment theory.
- Introduction to asset pricing theories.
- Style analysis and risk measures.
- Monte Carlo Simulation and Resampling.

# What this course is not:



- A course in economics, accounting or corporate finance.
- A "pure" statistics course.
- A course about financial markets or institutions.
- A survey of investments.
- A course on personal investments.
- A course on valuation.

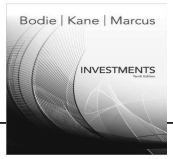


*API 141 Finance*
*Syllabus*
*Akash Deep*
*September 2, 2009*

**The typical intro to finance class**

| No. | Day | Date | Topic |
|---|---|---|---|
| 1 | Wed | Sep 2 | Introduction to finance and financial markets |
| 2 | Wed | Sep 9 | Present value and the opportunity cost of capital |
| 3 | Mon | Sep 14 | Valuing financial securities: Bonds |
| 4 | Wed | Sep 16 | Valuing financial securities: Equity |
| 5 | Mon | Sep 21 | Diversification, risk, and return measures |
| 6 | Wed | Sep 23 | Choosing a portfolio |
| 7 | Mon | Sep 28 | Case: *The State of South Carolina* |
| 8 | Wed | Sep 30 | The Capital Asset Pricing Model |
| 9 | Mon | Oct 5 | Case: *Communications Satellite Corporation* |
| 10 | Wed | Oct 7 | Efficient markets |
| 11 | Wed | Oct 14 | Arbitrage (Case: *Long Term Capital Management*) |
| 12 | Mon | Oct 19 | Risk management |
| 13 | Wed | Oct 21 | Forward and futures contracts |
| 14 | Mon | Oct 26 | Case: *Docier Industries* |
| 15 | Wed | Oct 28 | *Midterm Exam* |
| 16 | Mon | Nov 2 | Options |
| 17 | Wed | Nov 4 | Pricing of options |
| 18 | Mon | Nov 9 | Guest Speaker: Colin MacNaught, Assistant Treasurer for Debt Management, State of Massachuset |
| 19 | Mon | Nov 16 | Case: *BASIX* |
| 20 | Wed | Nov 18 | Real options |
| 21 | Mon | Nov 23 | Case: *Bidding for Antamina* |
| 22 | Wed | Nov 25 | Case: *Federal Deposit Insurance Corporation* |
| 23 | Mon | Nov 30 | Case: *Subprime Meltdown* |

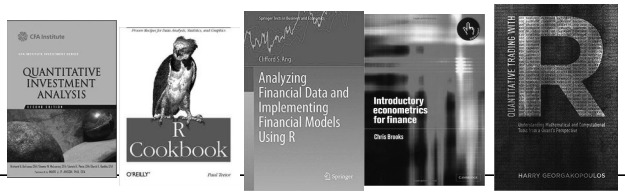
We will cover this material along with programming in R, simulation and more statistical rigor (among other things)




# Textbook



- There isn't one specific book; Hollis has electronic access to many useful books and we will link to them as needed.




# Aside: the BSAS and the CFA



B S A S
THE BOSTON SECURITY ANALYSTS SOCIETY, INC
*a Member Society of CFA Institute*

- A not-for-profit society for investment professionals in the Boston area
- Membership: 4,600
- Founding society of the CFA Institute
- www.bsas.org (student membership available).

## Chartered Financial Analyst®

The Economist

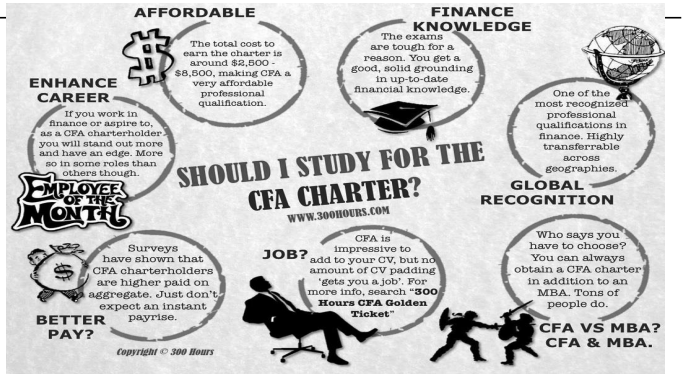
"Whereas there are tens of thousands of finance degrees available around the world, ranging from the excellent to the worthless, there is only one CFA, managed and examined by an American association of financial professionals, the CFA Institute. It used to be just an American qualification. But explosive growth has made it, in effect, global currency."

**Level I** – Knowledge and comprehension of fundamental investment tools and concepts

**Level II** – Application and analysis focused on asset valuation

**Level III** – Synthesis and evaluation of client needs and portfolio management issues

*Ethical and professional standards are emphasized at every level.*




See also http://www.lifeonthebuyside.com/mi-hedge-funds-vs-asset-management/



# Course Overview: Classes 1-5

- Introduction
- Get Comfortable with R
- Review basic statistics (descriptive, confidence intervals, hypothesis testing)

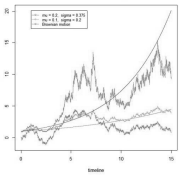

| Data | | 1 Sample | 2 Samples | >2 Samples |
|---|---|---|---|---|
| Continuous & Meets Normality Assumption | Means | $H_0$: $\mu = \mu_0$<br>$H_A$: $\mu \neq \mu_0$<br>One Sample T-Test | $H_0$: $\mu_1 = \mu_2$<br>$H_A$: $\mu_1 \neq \mu_2$<br>Two Sample T-Test | $H_0$: $\mu_1 = \mu_2 = \ldots = \mu_k$<br>$H_A$: At least one mean is different<br>Analysis of Variance |
| | Standard Deviations | $H_0$: $\sigma = \sigma_0$<br>$H_A$: $\sigma \neq \sigma_0$<br>Chi-Square Test | $H_0$: $\sigma_1 = \sigma_2$<br>$H_A$: $\sigma_1 \neq \sigma_2$<br>F-Test | $H_0$: $\sigma_1 = \sigma_2 = \ldots = \sigma_k$<br>$H_A$: At least one st. dev is different<br>Bartlett's Test |
| Continuous & Non Normal | Median | $H_0$: $\eta = \eta_0$<br>$H_A$: $\eta \neq \eta_0$<br>Wilcoxon Test | $H_0$: $\eta_1 = \eta_2$<br>$H_A$: $\eta_1 \neq \eta_2$<br>Mann-Whitney Test | $H_0$: $\eta_1 = \eta_2 = \ldots = \eta_k$<br>$H_A$: At least one mean is different<br>Kruskal-Wallis Test |
| Discrete | Proportion | $H_0$: $\Pi = \Pi_0$<br>$H_A$: $\Pi \neq \Pi_0$<br>Test for One Proportion | $H_0$: $\Pi_1 = \Pi_2$<br>$H_A$: $\Pi_1 \neq \Pi_2$<br>Test for Two Proportions | $H_0$: $\Pi_1 = \Pi_2 = \ldots = \Pi_k$<br>$H_A$: At least one proportion is different<br>Binomial Analysis of Means |

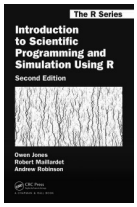
$H_0$ = Null Hypothesis
$H_A$ = Alternative Hypothesis
If $p <$ alpha risk, reject $H_0$.
If $p >$ alpha risk, do not reject $H_0$.



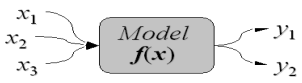
# Course Overview: Class 6

- The Bootstrap and Monte Carlo Simulation



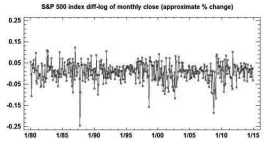

$$F \xrightarrow{\text{Random Sample}} (X_1^1, X_2^1 \ldots X_n^1) = \mathcal{X}_n^1 \quad \hat{\theta}_1$$
$$\vdots \qquad\qquad \vdots \quad \vdots$$
$$F \xrightarrow{\text{Random Sample}} (X_1^2, X_2^2 \ldots X_n^2) = \mathcal{X}_n^2 \quad \hat{\theta}_2$$
$$\vdots \qquad\qquad \vdots \quad \vdots$$
$$F \xrightarrow{\text{Random Sample}} (X_1^B, X_2^B \ldots X_n^B) = \mathcal{X}_n^B \quad \hat{\theta}_B$$

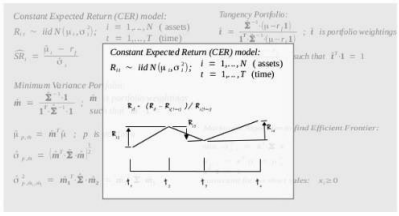





# Course Overview: Classes 7-8

- The Constant Expected Return Model

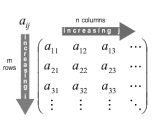


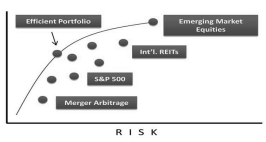



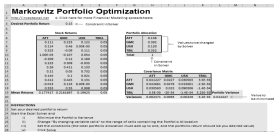



# Course Overview: Classes 9-11

- Matrices, Portfolio Models and Mean Variance optimization.





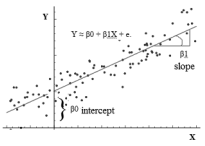

# Course Overview: Class 12

- Holiday

# Course Overview: Classes 13-15



- Simple Regression, the Market Model and Beta.





The Formula for the Beta Coefficient

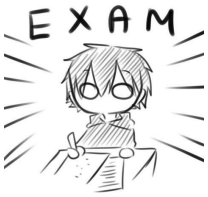
Beta is equal to the covariance of the returns of the stock with the returns of the market, divided by the variance of the returns of the market:

$$\beta_i = \frac{cov_{i,M}}{\sigma_M^2} = \frac{\rho_{i,M}\sigma_i}{\sigma_M}$$

# Course Overview: Class 16



- Midterm




# Course Overview: Classes 17-19

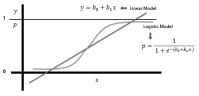


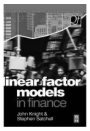
- Multiple Regression, diagnostics, Fama-French and logistic regression.

**Multifactor Pricing Models**

$$R_{jt} = a_j + \beta_{j1}F_{1t} + \beta_{j2}F_{2t} + \ldots + \beta_{jk}F_{kt} + \varepsilon_{jt}$$

$R_{jt}$ = return on asset $j$ during a specified time period
$E_j$ = expected return for asset $j$
$\beta_{jk}$ = reaction in asset $j$'s returns to movements in a common factor
$F_k$ = a common factor influences the returns on all assets
$\varepsilon_{jt}$ = a unique effect on asset $j$'s return that, by assumption, is completely diversifiable in large portfolios and has a mean of zero

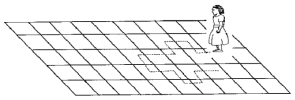


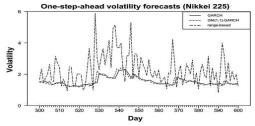




# Course Overview: Classes 20-23



- Time Series Topics-Random Walks, ARCH, GARCH, Cointegration and Pairs Trading



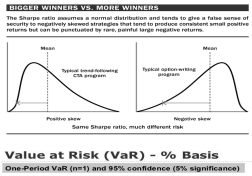


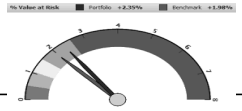


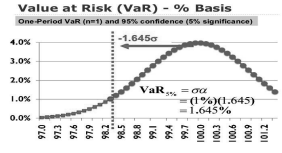
# Course Overview: Class 24



- Risk Measures and Value at Risk


# Course Overview: Class 24



- Discuss Projects

# Course Overview: Class 25



- Holiday


http://www.forbes.com/sites/steveschaefer/2015/06/17/new-money-masters-fidelity-chuck-myers-small-caps/#153331f07582

# Course Overview: Class 26



- Guest Speaker: Chuck Myers

Fidelity INVESTMENTS

40 Under 40 - Fortune
fortune.com/40-under-40/2013/

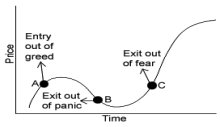
| | |
|---|---|
| Age | 37 |
| Title | Portfolio Manager |
| Company | Fidelity Investments |

Myers is a bright star at Fidelity, where he manages $10.5 billion in small-cap stocks. You've probably never heard of his top holdings — companies like Berry Petroleum and j2 Global). But his picks during the financial crisis have driven his two mutual funds to outperform all but a handful of competitors over the past five years. His secret? Avoiding distractions. He doesn't have voicemail and spends most days inside his quiet office studying dozens of companies. Advice he would have given his 20-year-old self: Buy Apple.

# Course Overview: Class 27



- Behavioral Finance and Wrap Up





Wrap-Up Time to this Course

# A Theory of Risk



- A lot of this course will consist of using statistics to analyze and understand stock returns.
- How to model returns, and understand their behavior and risk.
- This is important since quite interesting things can happen in the stock market.

# December 27, 1999



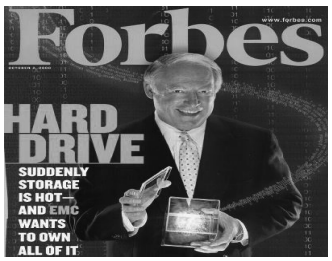

# After Cover….



- Went from $100 to $6

# EMC



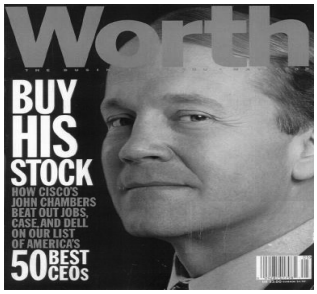



**EMC (10/02/00) = 97.36**

**EMC (09/26/02) = 5.70**

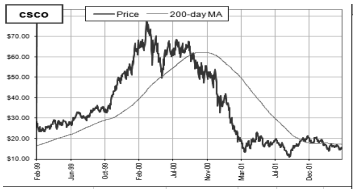

# CSCO






**CSCO (05/15/00) = 60.00**

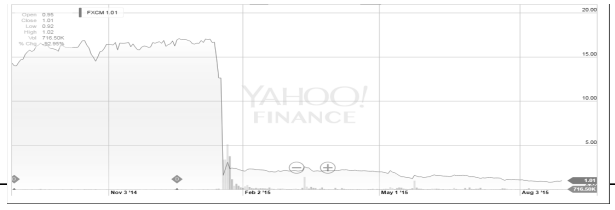
**CSCO (09/26/02) = 11.94**




# More Recently



- Brokerage FXCM Inc., stock dropped 90% in one day reacting on losses of the client in forex trades.




# More Recently (in mid August)



**StemCells Inc. (STEM)** ☆ Add to watchlist

NasdaqCM - NasdaqCM Real Time Price. Currency in USD

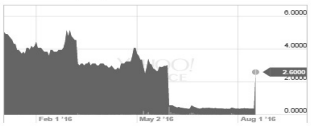
2.60 +2.2278 (+598.55%) 1.90 -0.70 (-26.92%)

At close: 4:00 PM EDT After hours: 5:50PM EDT

Summary | Conversations | Statistics | Profile | Financials | Options | Holders | Historical Data | Analysts

| | | | |
|---|---|---|---|
| Open | 0.45 | Market Cap | 30.16M |
| Prev Close | 0.37 | P/E Ratio (ttm) | -0.64 |
| Bid | 2.06 x 100 | Beta | -0.84 |
| Ask | 2.10 x 4200 | Volume | 58,370,502 |
| Day's Range | 0.42 - 2.99 | Avg Vol (3m) | 616,051 |
| 52wk Range | 0.33 - 7.56 | Dividend & Yield | N/A (N/A) |
| 1y Target Est | 22.00 | Earnings Date | May 6, 2016 - May 9, 2016 |

1D 5D 1M 6M YTD 1Y 2Y 5Y 10Y MAX Interactive chart



# We'll learn how to study risk and return



Mutual Fund Comparison

Tickers (limit 5): VFINX VBAIX FCNTX FSCRX Show Funds

Compare: Returns Risk Fees Holdings

| Symbol | Fund Name | 1 Wk | 13 Wk | YTD | 1 Yr | 3 Yr (Annualized) | 5 Yr (Annualized) | 10 Yr (Annualized) |
|---|---|---|---|---|---|---|---|---|
| VFINX | Vanguard 500 Index;Inv | -0.66% | 3.83% | 7.57% | 14.11% | 11.56% | 15.30% | 7.45% |
| VBAIX | Vanguard Bal Index;Inst | -0.39% | 3.54% | 7.26% | 10.46% | 8.47% | 10.48% | 7.03% |
| FCNTX | Fidelity Contrafund | -0.29% | 3.15% | 3.58% | 9.14% | 11.42% | 14.36% | 8.78% |
| FSCRX | Fidelity Sm Cap Disc | -0.59% | 2.53% | 8.93% | 7.34% | 6.45% | 15.80% | 11.18% |

# More Statistics



Mutual Fund Comparison

Tickers (limit 5): VFINX VBAIX FCNTX FSCRX Show Funds

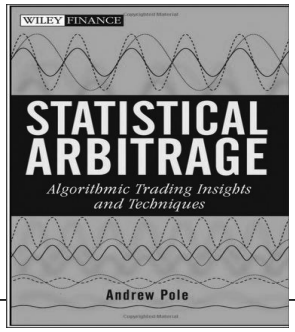
Compare: Returns Risk Fees Holdings

(All calculations based on a 3-yr period. See below for comparisons)

| Symbol | Fund Name | Alpha | Beta | RSquared | Std. Dev. |
|---|---|---|---|---|---|
| VFINX | Vanguard 500 Index;Inv | -0.01 | 1.00 | 1.00 | 3.16 |
| VBAIX | Vanguard Bal Index;Inst | 0.12 | 0.60 | 0.95 | 1.93 |
| FCNTX | Fidelity Contrafund | 0.07 | 0.94 | 0.86 | 3.19 |
| FSCRX | Fidelity Sm Cap Disc | -0.34 | 0.98 | 0.67 | 3.75 |

# Many Ideas Under a Fancy Umbrella






# Stat Example: Pairs Trading

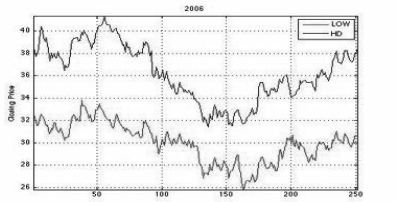


- Pairs trading is also known as statistical arbitrage, and is a very interesting concept.
- In simple terms, one looks for stocks that seem to move together, like GOOG and AAPL, or HD and LOW.
- When their historical relationship gets out of whack, one shorts the overbought one and buys the oversold one.
- It's a market neutral concept, as one invests the same amount of money in each position.

# Example with $HD and $LOW



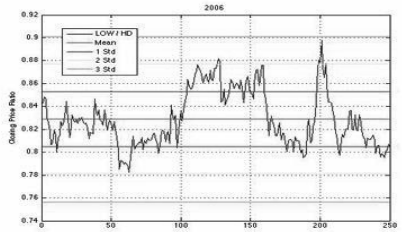
- Home Depot and Lowes seem to move in similar fashion




# Plot the Ratio $LOW/$HD



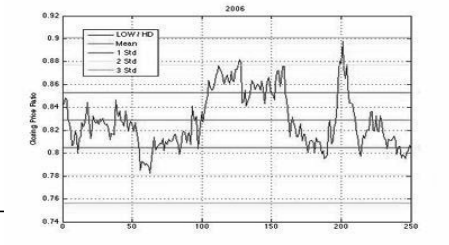
- We are interested in the ratio of prices




# Examine the mean and std deviations.



We see a price ratio (blue curve) that appears to oscillate around a mean (purple line). And I've also added the 1-, 2-, and 3-standard deviation lines just to get a feel for how far away from the mean the oscillations might go.




# Set a trap



- We only want to trade when we have a very good edge in our favor.
- We see that in 2006 the pair ratio often makes excursions 1 or 2 standard deviations above or below the mean. We could spring the trap then but we might end up making lots of trades for tiny profits, and the commissions could eat us alive too.

# The Best Trap



- Examining the figure, we do see **one time** when the ratio appeared to go almost 3 standard deviations from the mean. Therefore that will be our criteria for 2007.
- **If we see an excursion +/- 3 standard deviations from the mean, we'll enter a pair trade. And whenever the ratio returns to the mean, we'll exit both positions.**

# Pairs Trading is Popular



- Pairs trading is an effective, relatively low cost strategy that is now well know.
- People use the price difference of the two stocks, or the ratio of the two stocks….there are many variations (see web site for some papers).

# Info about R and pairs trading



Introduction to Cointegration and Pairs Trading

April 15, 2011

By Edwin Chen

Introduction Suppose you see two drunks (i.e., two random walks wandering around. The drunks know each other (they're independent), so there's no meaningful relationship between their paths. But suppose instead you have a drunk walking with her dog.

Seasonal pair trading

January 10, 2011

By Dzidorius Martinaitis

quanttrader.info is a good quantitative repository, where I found an idea abo seasonal spreads play. The idea of seasonal pair trading differs from pair trading in a way, that it doesn't try to find deviation from the spread's mean read pattern to find...

QuantTrader

.. quantitative trading, programming and more

Subscribe to feed Blog Pair Trading

Pair-Trading with S&P500 Companies – Part II.

April 16, 2011

By QuantTrader

Today I'm going to share with you further outcomes of my research in statistical arbitrage trading technique - pair-trading. In the first part of pair-trading with S&P500 Companies I used downloaded data from yahoo to identify co-integrated pairs. …

Poor man´s pairs trading...

April 11, 2010

By M. Parzakonis

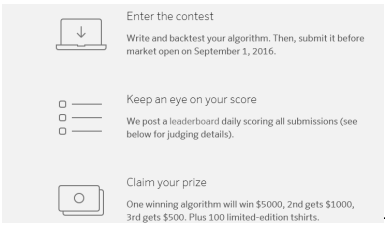
There is a central notion in Time Series Econometrics, cointegration. Loosely it refers to finding the long run equilibrium of two non-stationary series. As the most know non-stationary series examples comes from finance, cointegration is nowadays a tool for traders (not a common one though!). They use it as the theory behind pairs trading…

**http://www.r-bloggers.com/**

# Algorithmic Trading



- Quantopian (located in Kendall Square)




1. You use our platform to create investment algorithms.

2. We evaluate your algorithms, and selected authors receive an offer to license their algorithm.

3. When a selected algorithm generates positive returns, the author gets a cut.

Our job is to raise the capital, handle all day-to-day trading operations, provide useful data sets, and build the best platform in the world for creating investment strategies.

# Movie Time-Algorithms



- Kevin Slavin, How Algorithms Shape our World.

As an entrepreneur, Kevin has successfully navigated and integrated the areas of gaming, new media, technology, and design. As Co-founder of Area/Code in 2005, Kevin was a pioneer in rethinking game design and development around new technologies (like GPS) and new platforms (like Facebook). Area/Code worked to develop next-generation game experiences not only for major consumer product groups like Nokia, Nike and Puma but for media giants such as MTV, Discovery Channel, CBS and Disney. Their Facebook game Parking Wars, commissioned by A&E Television to promote its show of the same name, served over 1 billion pages in 2008. The company was acquired by Zynga in 2011, becoming Zynga New York.

# Computing In a Nutshell



| Name | Advantages | Disadvantages | Open source? | Typical users |
|---|---|---|---|---|
| R | Library support; visualization | Steep learning curve | Yes | Finance; Statistics |
| Matlab | Elegant matrix support; visualization | Expensive; incomplete statistics support | No | Engineering |
| SciPy/NumPy /Matplotlib | Python (general-purpose programming language) | Immature | Yes | Engineering |
| Excel | Easy; visual; flexible | Large datasets | No | Business |
| SAS | Large datasets | Expensive; outdated programming language | No | Business; Government |
| Stata | Easy statistical analysis | | No | Science |
| SPSS | Like Stata but more expensive and worse | | | |

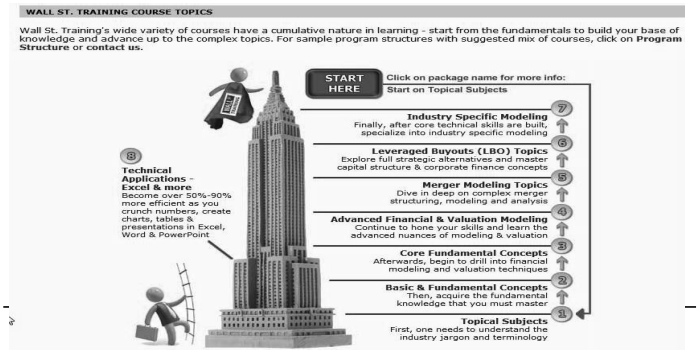
# General Computing Information



- This course will use slightly Excel but mostly R
- We will start to go over these two packages today with an emphasis on R.
- Excel is good for data management and a graphical user interface (sometimes called a GUI).
- R is better for flexibility and computational power.

# Knowing Excel is a Necessary Evil






# Some not so esoteric Excel ideas



- We might see these ideas (or not) but you should pick them up sometime (not for this class but for future employment):
  - Conditional formatting
  - Buttons and spinners
  - Pivot tables
  - Making graphs
- There is a nice basic intro to Excel located at http://cameron.econ.ucdavis.edu/excel/excel.html (linked on the class web site).

# Introduction to R




# What is R?



- A computer language, with orientation toward statistical applications
- Relatively new
- Growing rapidly in use
- Its what the professionals are now using

# Some R commands (details later)



```
> getSymbols("AAPL")
> chartSeries(adjustOHLC(AAPL,
  use.Adjusted=TRUE),theme="white")
> addSMA(50,col=2)
> addSMA(200,col=3)
```

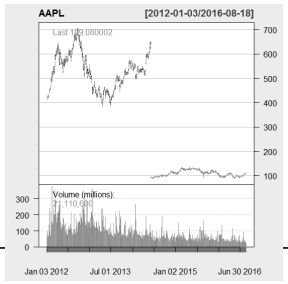
# The Resulting Graph




# What are "adjusted" prices??



- `chartSeries(AAPL,theme="white")`




# R is a tool for...



**Data Manipulation**
- connecting to data sources
- slicing & dicing data

**Modeling & Computation**
- statistical modeling
- numerical simulation

**Data Visualization**
- visualizing fit of models
- composing statistical graphics

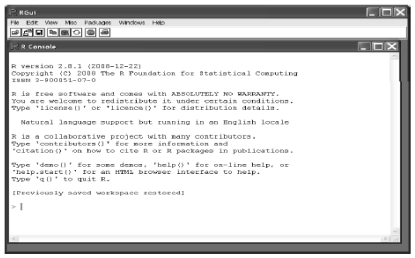
munge
model
visualize

# R is an environment



# Its interface is plain






# R's Ups and Downs



- Plusses
  - Completely free, just download from Internet
  - Runs on many operating systems
  - Many add-on packages for specialized uses
  - Open source
- Minuses
  - Obscure terms, intimidating manuals, odd symbols, inelegant output (except graphics)

## Installing R



- www.r-project.org/
- download from CRAN
- select a download site
- download the base package at a minimum
- download contributed packages as needed
- Go to web site for detailed instructions!

## To Do




- Install R
- See website for introduction to R materials (for example datacamp)
- Section starts next week [no sectioning-go to as few or as many as needed-details on course web site]
- **Homework out Monday due the following Monday**