

Applied Analytics: Frameworks and Methods 2

Introduction

No Phone

No Photographing

No Audio Recording

No Video Recording

Lecture Content and Materials Should Not Be Posted or Shared

Online or Offline Without Explicit Permission

Time and Location

Day/Time: Tuesday, 8:10am-10:00am

Modality: ON-CAMPUS

Location: 417 Mathematics Building

- 1/18 and 1/25 (lecture 1 and 2)
 - classes meet online
 - use zoom link under "Zoom Class Session" page on canvas
- Starting 2/1 (lecture 3 and all lectures follow)
 - classes meet on campus
 - 417 Mathematics Building

Teaching Team - Professor

- Kitty Kay Chan
 - Email Address: kkc2139@columbia.edu
 - Office Hours: Tuesdays, 10:00am – 12:00am
 - Location:
 - 1/18 and 1/25 (lecture 1 and 2)
 - office hours meet online
 - use zoom link under "Zoom Class Session" page on canvas
 - Starting 2/1 (lecture 3 and all lectures follow)
 - Office hours meet on campus
 - Lewisohn Hall 502F

Teaching Team -Associates

- Elena Dubova

- Address: ed28o1@columbia.edu
- Office Hours: Wednesdays, 5:00-6:00 pm
- Zoom link to join:

<https://columbiauniversity.zoom.us/j/92823692906?pwd=doxjRHNrNEJoVDNXYo5iS1FEeDg3Zz09>

- Nicholas Porter

- Email Address: nicholas.porter@columbia.edu
- Office Hours: Mondays, 11:00-12:00 pm
- Zoom link to join:

<https://columbiauniversity.zoom.us/my/nbp2119>

Outline

- Framework and Methods – 2
- About the Course
- Project
- Getting Ready

Framework and Methods - 2

- Traditional data: E.g., transactional data and survey data
- People generating Data
 - Email, Social Media, ...
 - Personal devices such as Fitbit, smartphone, Nest, iRobot, Nike
- Other Devices
 - IoT: On board computers on machines from cars to airplanes
- Democratization of data
 - Open Data revolution is afoot led by governments. Data.gov shares over 300,000 datasets (as of 2018)
 - Websites share data for free directly from website or through an API. E.g., Google Trends, Yahoo Finance, Twitter

Data

- Governments and companies have made it easier for people to share data with them.
 - Software and hardware diagnostics
 - Reporting potholes, graffiti, crime
 - Service complaints
- Harvesting of data
 - Previously unused data is being captured. E.g., Electronic Health Records
 - Speech, pictures, video
 - Web scraping

Analytics

Fundamental Forces Changing *Everything*



- Sequel to Framework and Method – 1
- Unsupervised Learning
- Unstructured Data
- Time Series
- Deep Learning
- Spatial Data

Supervised and Unsupervised Learning

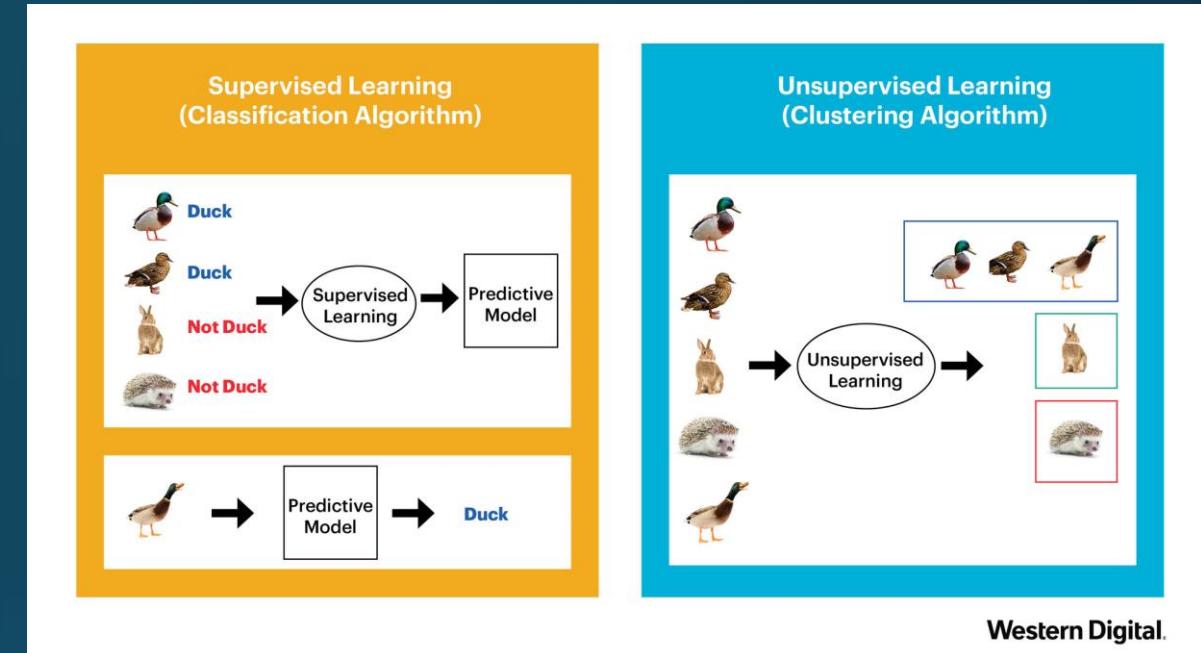
■ Supervised Learning

- We have data on both predictors and outcome.
- Also, known as labeled data
- E.g., regression, trees
- Examined in Framework and Methods 1

■ Unsupervised Learning

- There is data on a set of variables but no associated outcome or response variable.
- E.g., cluster analysis, factor analysis, market basket analysis

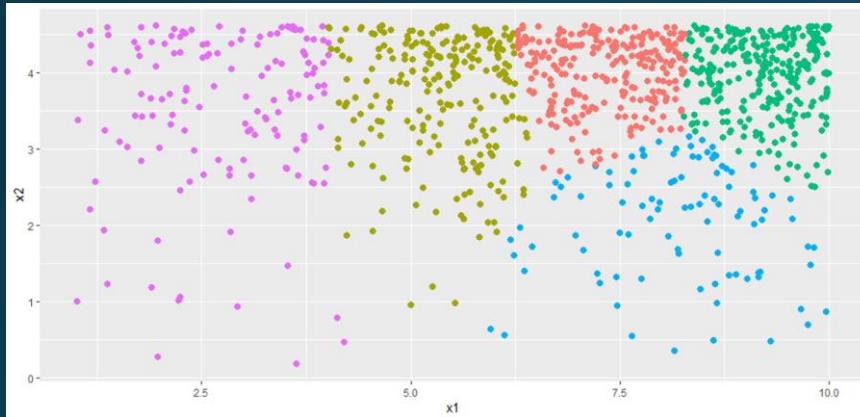
■ This course will look at both supervised Learning and unsupervised learning



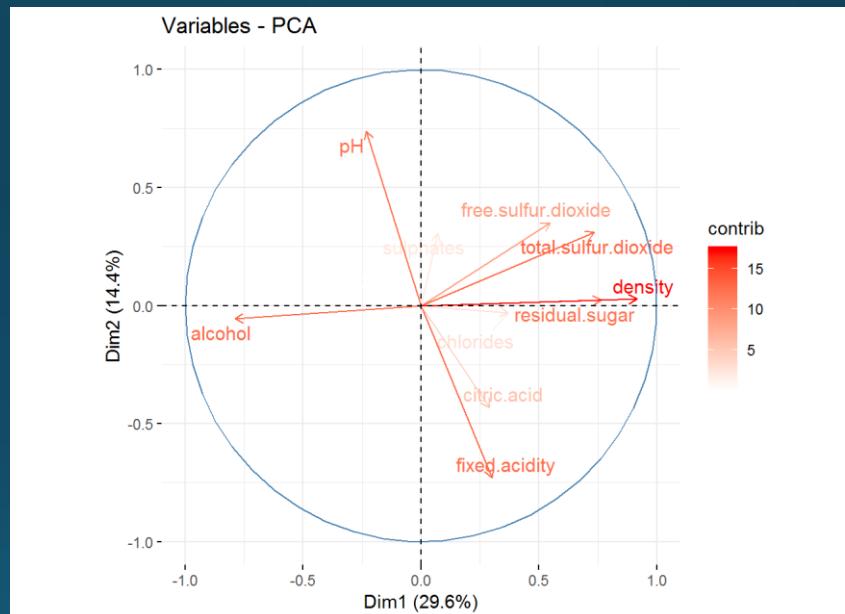
Western Digital.

Supervised and Unsupervised Learning

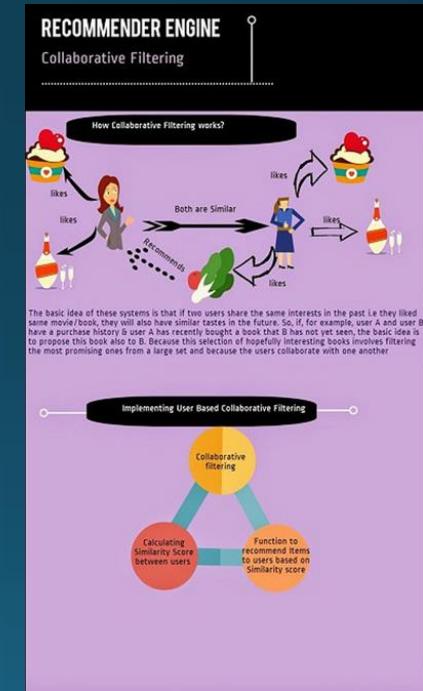
Clustering

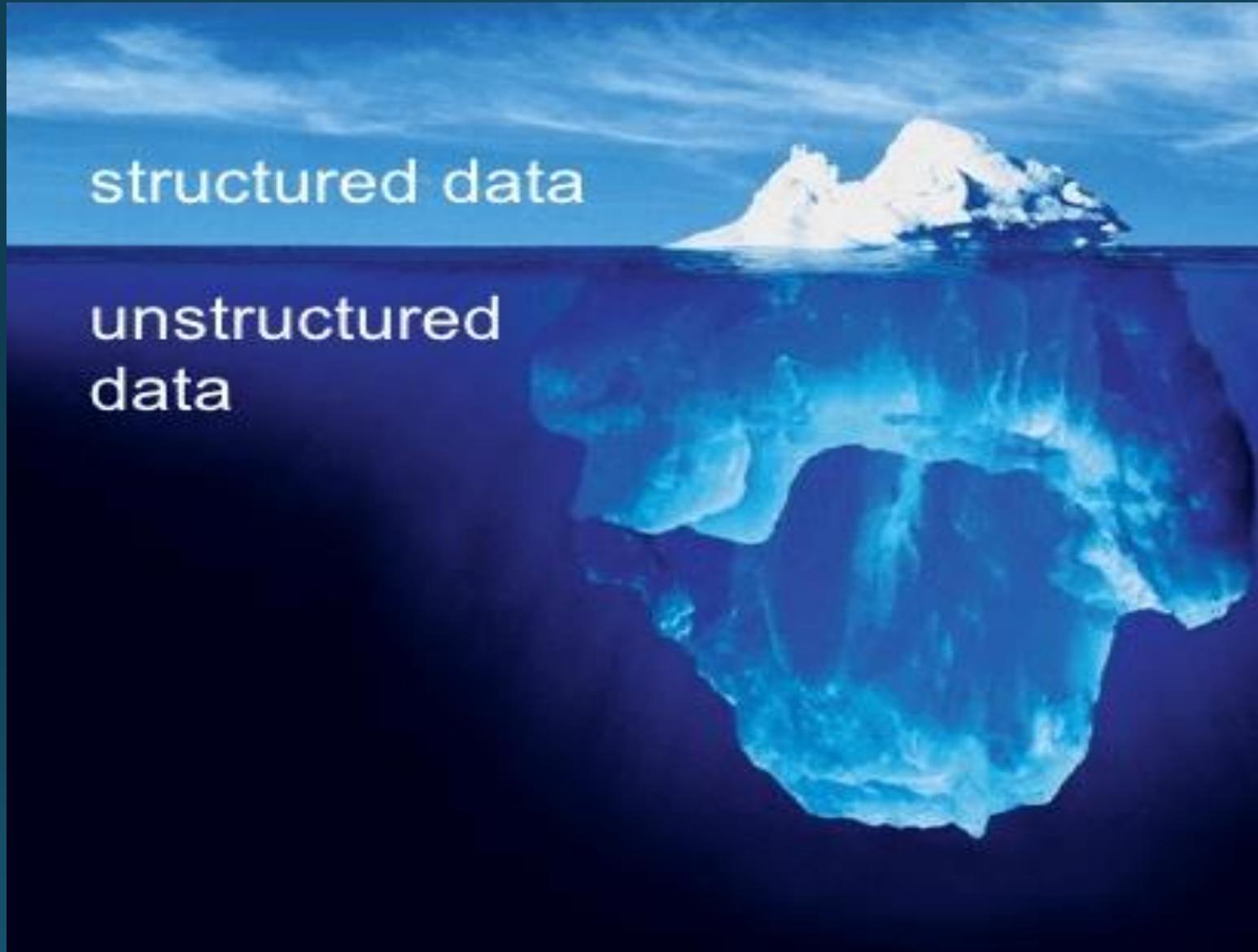


Dimension Reduction



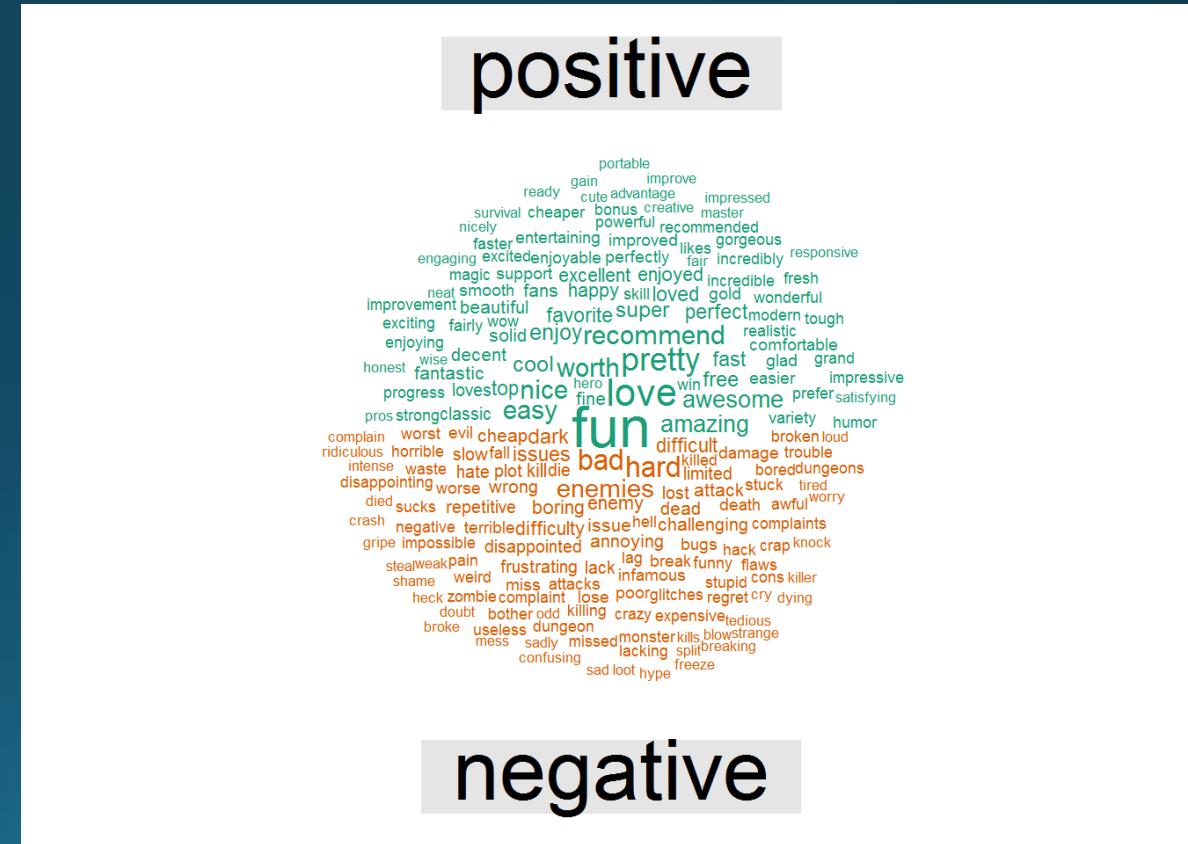
Associations and Recommendations





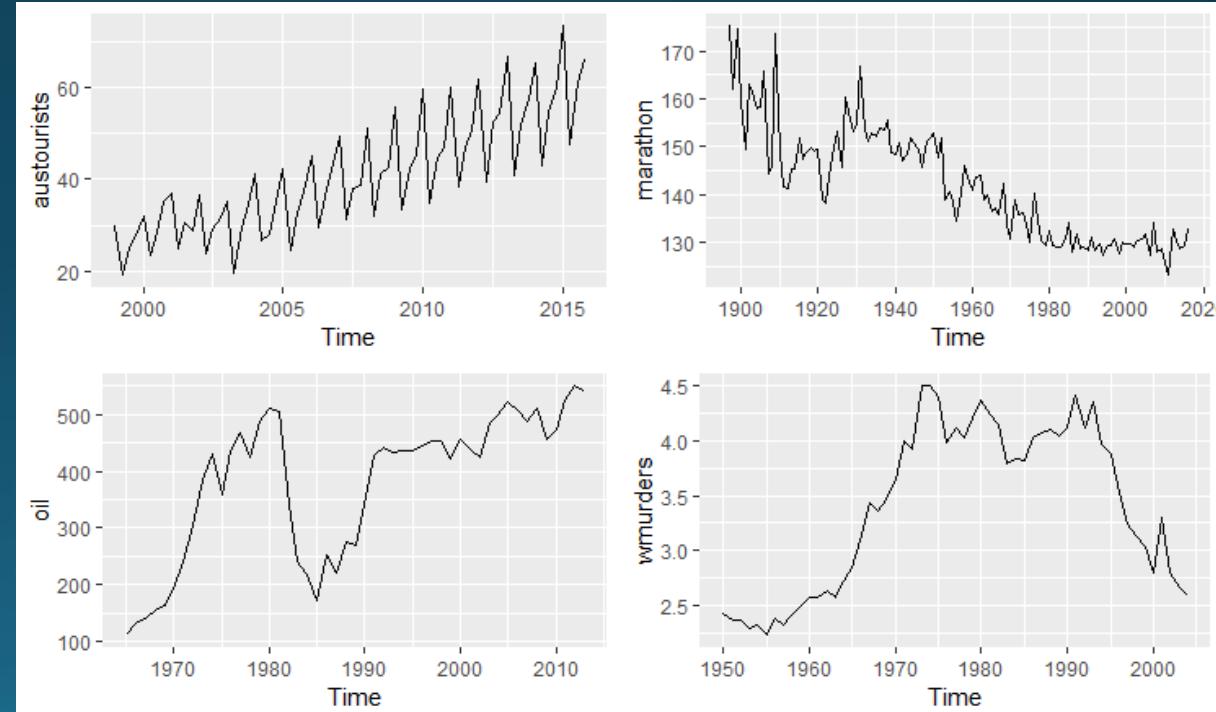
Structured vs. Unstructured Data

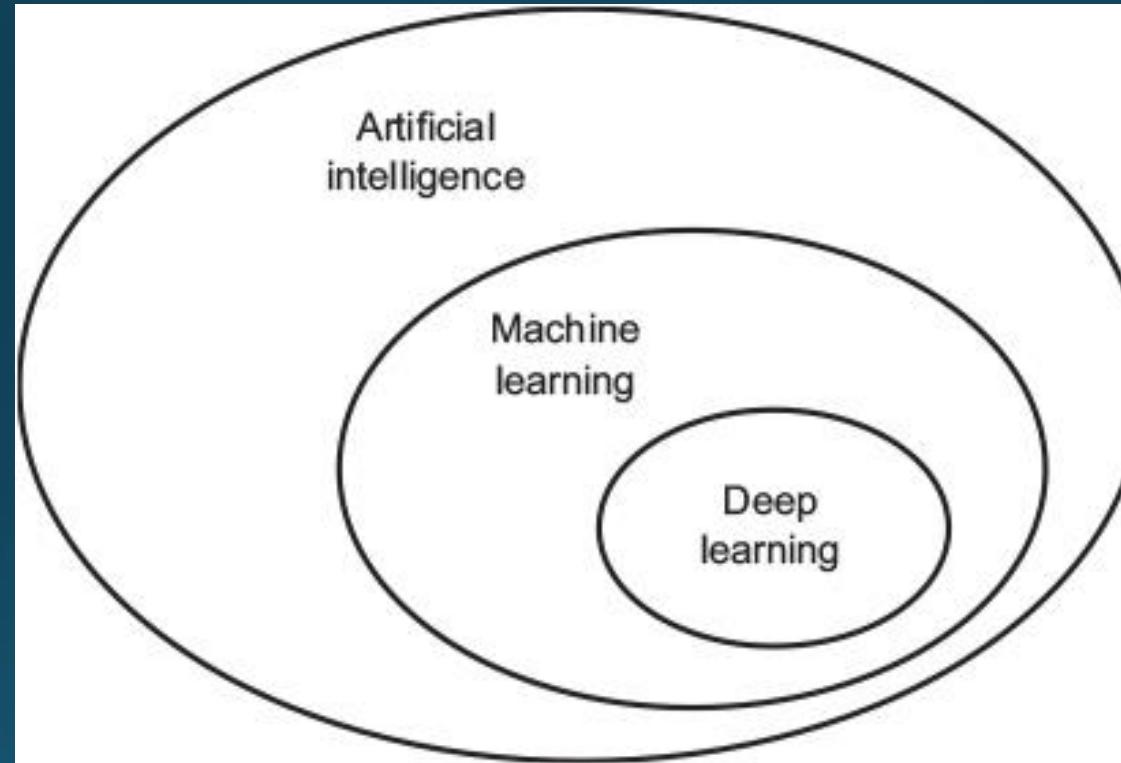
- Structured Data: Numerical data that fits neatly into rows and columns.
- Unstructured data
 - Words
 - Images
 - Video



Cross-Sectional vs. Time Series

- Cross-sectional: Data gathered at a single point in time
- Longitudinal or time series: Data gathered over a span of time

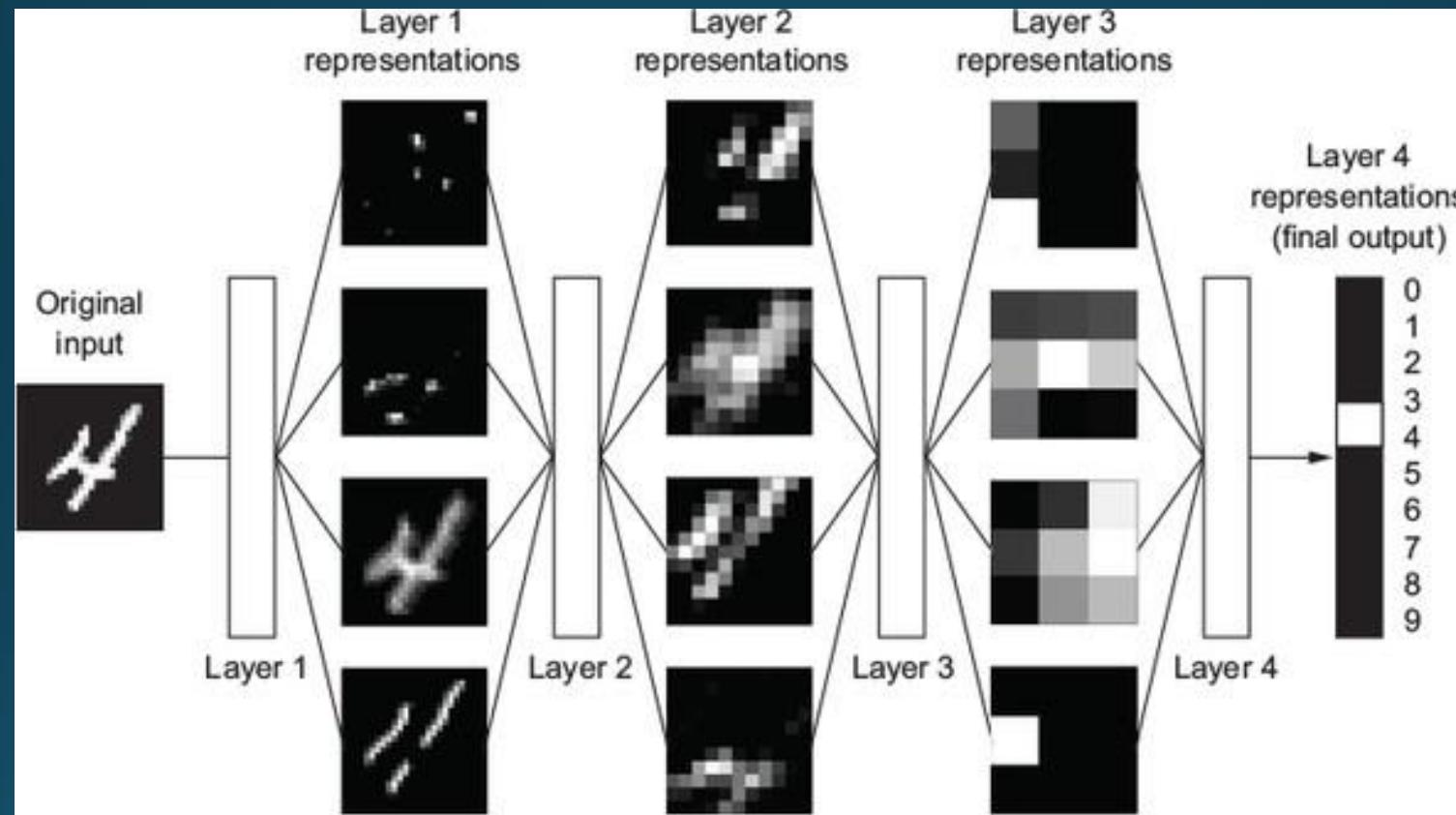




Deep Learning

- Machine Learning:
 - Outcomes are predicted from Predictors.
 - The mapping of predictors to outcomes is learned from data

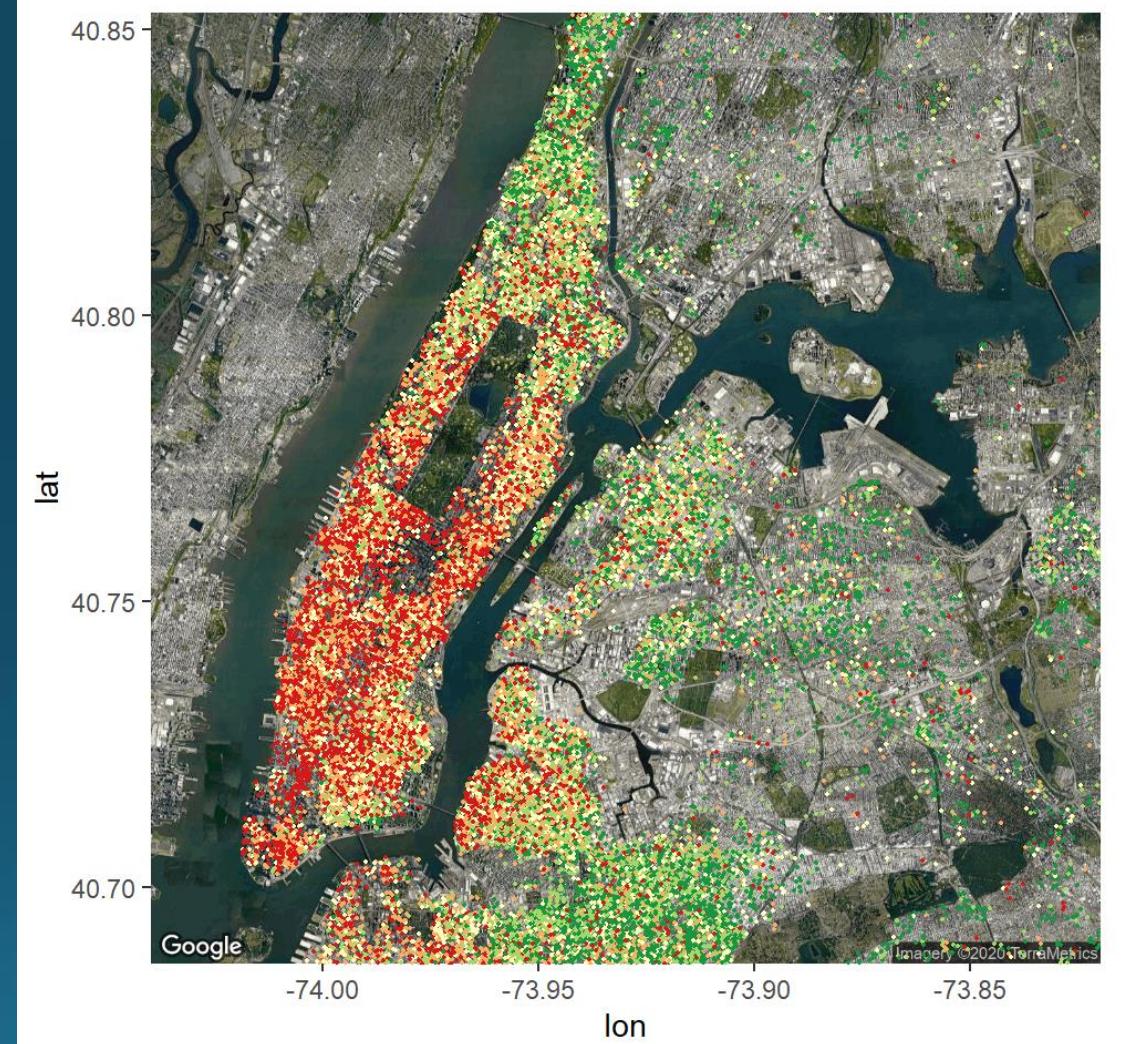
- Deep Learning:
 - The mapping of predictors to outcomes involves multiple layers
 - These layered representations are learned via models called neural networks, structured in literal layers stacked on top of each other.



Neural Network framework for identifying a digit

Spatial Data

- Data about location
- You are where you live



About the Course

About the Course

- Overview
- Learning Objectives
- Readings
- Resources
- Assessment
- Grading
- Policies
- Schedule

About the Course - Overview

- Course will
 - Identify and leverage analytical techniques to address organizational problems
 - Construct analytical models to generate actionable insights from structured and unstructured data
 - Implement analytical techniques using R

- Approach of the course is to help you to “*learn by doing*”
 - Discuss analytical technique at high level
 - Implement on a dataset in class
 - Hands-on assignment to tackle a problem
 - Exam and project to demonstrate learning

About the Course - Readings

- R for Marketing Research and Analytics by Chris Chapman, Elea McDonnell Feit (Springer)
(referred to as Analytics Text)
- Building a Recommendation System with R by Suresh K. Gorakala, Packt Publishing
(referred to as Recommendation Text)
- Text Mining with R by David Robinson and Julia Silge (O'Reilly Media)
(referred to as Text Mining Text)
- Forecasting: Principles and Practice by Rob J Hyndman and George Athanasopoulos (Otexts)
(referred to as Forecasting Text)
- Deep Learning with R by Francois Chollet and J. J. Allaire (Manning Publications)
(referred to as Deep Learning Text)
- Applied Spatial Data Analysis with R by Roger S. Bivand, Edzer Pebesma and Virgilio Gómez-Rubio (Springer)
(referred to as Spatial Text)

Resources

- Columbia University Information Technology
 - University Provided discounted software downloads
- Columbia University Library
- SPS Academic Resources

Assessment

- Assignments: 20% (Four assignments, each worth 5%)
- Project: 30%
- Final Exam: 40%
- Class Participation 10%

Grading Scale

Grade	Percentage
A+	98–100 %
A	93–97.9 %
A-	90–92.9 %
B+	87–89.9 %
B	83–86.9 %
B-	80–82.9 %
C+	77–79.9 %
C	73–76.9 %
C-	70–72.9 %
D	60–69.9 %
F	59.9% and below

- Course Policies

- Participation and Attendance
- Late Work
- Citation and Submission

- School Policies

- Copyright policy
- Academic Integrity
- Accessibility

- Participation and Attendance

You are expected to attend all class sessions.

- Late Work

No credit is granted to any written assignment submitted after the due date. Late assignments must receive prior permission from the instructor, and a penalty will be assessed. All written assignments must be submitted to the course website; email submissions or re-submissions are not accepted.

- Citation & Submission

All written assignments must use APA citation, cite sources, and be submitted to the course website (not via email).

Policies – School Policies

■ Copyright Policy

Please note—Due to copyright restrictions, online access to this material is limited to instructors and students currently registered for this course. Please be advised that by clicking the link to the electronic materials in this course, you have read and accept the following:

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

- Academic Integrity

Columbia University expects its students to act with honesty and propriety at all times and to respect the rights of others. It is fundamental University policy that academic dishonesty in any guise or personal conduct of any sort that disrupts the life of the University or denigrates or endangers members of the University community is unacceptable and will be dealt with severely. It is essential to the academic integrity and vitality of this community that individuals do their own work and properly acknowledge the circumstances, ideas, sources, and assistance upon which that work is based. Academic honesty in class assignments and exams is expected of all students at all times.

SPS holds each member of its community responsible for understanding and abiding by the [SPS Academic Integrity and Community Standards](#). You are required to read these standards within the first few days of class. Ignorance of the School's policy concerning academic dishonesty shall not be a defense in any disciplinary proceedings.

- Accessibility

Columbia is committed to providing equal access to qualified students with documented disabilities. A student's disability status and reasonable accommodations are individually determined based upon disability documentation and related information gathered through the intake process. For more information regarding this service, please visit the [University's Health Services website.](#)

Schedule

- View detailed schedule on Canvas

Module	Lecture Date	Topic	Assignment Opens (at 10 am, Tuesday)	Activity/Assignment Due (at 11:59pm, Monday)
1	1/18/2022	Introduction		
2	1/25/2022	Basic Clustering		
3	2/1/2022	Advance Clustering		
		Assignment 1	2/1/2022	2/14/2022
4	2/8/2022	Dimension Reduction		
5	2/15/2022	Text Mining Part 1		
6	2/22/2022	Text Mining Part 2		
		Assignment 2	2/22/2022	3/7/2022
		Project Proposal		2/28/2022
7	3/1/2022	Association Rules		
8	3/8/2022	Recommender System		
Holiday	3/15/2022			
		Assignment 3	3/8/2022	3/21/2022
9	3/22/2022	Time Series		
		Assignment 4	3/22/2022	4/4/2022
10	3/29/2022	Neural Networks		
11	4/5/2022	Spatial Analysis		
12	4/12/2022	Guest Speaker		
		Project Presentation Deck		4/18/2021 (at 12 noon)
13	4/19/2022	Presentation on Project		
		Project Final Report		4/19/2021 (at 11:59 pm)
14	4/26/2022	Final Exam		

Assignments (20%)

- Assignment 1: Clustering (5%)
- Assignment 2: Text Mining (5%)
- Assignment 3: Association Rules and Recommender Systems (5%)
- Assignment 4: Time Series Analysis (5%)

Project

- A key feature of this class is applying what you have learnt in the curriculum to a real-world dataset.
- The project requires you to work in a group (3-5 members) to perform all of the steps of a typical data analysis project over the course of the semester.
- Your report is expected to be independently developed by your team.
- You are free to select a problem, relevant data, and the types of analyses that you will perform.
- We encourage you to apply your own judgments in deciding how to frame the problems, structure the analysis, and select appropriate methods.
- This project should be something more than repeating the homework assignments on a different data set; you should aim to make a comprehensive and compelling analysis of an important problem in a domain or application of your choice.

DISCLAIMER

STUDENTS TAKING PART IN THIS PROJECT MUST REVIEW AND UNDERSTAND THE GUIDELINES AND LIMITATIONS ASSOCIATED WITH THE TERMS AND CONDITIONS OF THE COMPANY.

Project - Three graded components

- Project Proposal (5%)

Identify research problem, find data, and prepare data for analysis.

- Final Report (20%)

A comprehensive report of analysis, findings, and recommendations.

- Presentation (5%)

A group presentation highlighting the research problem, analysis conducted and conclusions.

Project Proposal - Details

- Identify a research problem in a domain that the group is interested in.

You may have certain beliefs you may want to put to the test or you may find inspiration for the problem in recent news events. For instance, you may be interested in examining the influence of social media activity on stock price. The more interesting the problem, the more exciting it will be for your group to work on and will make for a more interesting final presentation. Since this is a team effort, ensure everyone on the group buys in to the idea.

- Construct a set of research questions or hypotheses that you would like to answer.

Examine published research in the domain or on the subject chosen to select questions that have not been studied or have received only limited attention. Including citations to published articles reviewed is recommended.

Project Proposal - Details

- Next, look for data that promises to address the research problem.
 - The information must come from publicly available sources that can be verified and do not require the permission of a third party.
 - The dataset may be one that has been curated and ready to download or one that requires some active effort such as gathering web data through an API client (e.g., social media activity from Twitter, or financial data from Yahoo Finance), directly linking to an API, or capturing web page data. Alternatively, the information may be spread over multiple datasets that may need to be integrated before they can be used.
 - While any source of publicly available data is acceptable, keep in mind that you can develop a more original and unique project when using data from less studied sources. (For instance, nearly every data set from kaggle.com has multiple examples of fully developed projects with code available on publicly available websites.) While you are not prohibited from selecting such a project, you are encouraged to look more widely to select a good source of information. This will ultimately be helpful for your own learning and professional development.

Project Proposal - Details

- Find a suitable dataset
 - greater than 1000 observations or more than 8 variables
- Examine description of the variables in the dataset.
- Explore the data by examining descriptive summaries and visualizations, and investigate anomalies.
- Next, prepare the data for analysis which may include imputing missing values, recoding variables, creating new variables, extracting new features, and restructuring the data.
- While analysis is not a part of this phase of the Project, actively think of ways in which the data may be analyzed to address the research questions.

Project Proposal – Assessment

<u>Criteria</u>	<u>Points</u>
Raw dataset meets requirements for Project	20
Properly described data and variables	20
Identified and resolved key issues with raw data	20
Effectively demonstrated the use of functions in R to clean the data	20
Clearly articulated question(s) to be answered using the data	20

Project Proposal - Submission

What to submit:

Submit the following as part of the Project Proposal:

1. Data: Submit the raw dataset. If the dataset is too large, post the dataset to a cloud drive and share a persistent link. You should not expect the teaching staff to create accounts or log in to websites to download the data.
2. R Code: Share R code that can be used to obtain the cleaned dataset from the raw dataset. The R code may be submitted as an R script (.R), R Markdown (.rmd), or a knit R Markdown (.html) file.
3. Cleaned dataset: Submit the cleaned dataset obtained after processing the raw dataset. If the dataset is too large, share a persistent link to a cloud source.

Project Proposal - Submission

4. Written progress report (text, Word, PDF, or HTML format)
(2-3 written pages, no limits on spacing and font size):

The write-up should include:

- description of the research problem
- outline research question to be examine using the data
- a short narrative on the dataset
- a brief description of each variable
- a summary of the data processing steps undertaken

Project Proposal – Due February 28th at 11:59 pm

How to submit:

- Only one member of your group need submit this assignment on behalf of the group.
- To complete your submission:
 - i. Click the blue Submit Assignment button at the top of the Project Proposal page under module 6.
 - ii. Click the Choose File button, and locate your submission. Repeat for each file you submit.
 - iii. Feel free to include a comment with your submission.
 - iv. Finally, click the blue Submit Assignment button.

(depending on canvas web design update, this process may change)

- Select one or more of the analytical techniques covered in this class and apply it to the clean data to address the research questions constructed in the Project Proposal . Interpret results from analysis and express conclusions in a form that is understandable to decision makers.

1. Select Analytical Technique:

- This course covers a number of analytical techniques ranging from clustering to text mining to spatial analysis.
- You must select at least one technique from this class to address the problem outlined in Deliverable 1.
- Of course, you may use more than one technique from this class and you may also combine it with techniques covered in Frameworks and Methods 1.

Project: Final Report – Some Steps to Consider

2. Focus and prioritize your intended plans for analysis
 - Incorporate the feedback you receive from the teaching staff on the Proposal .
3. Build a draft of the initial analyses
 - Draft description of analytical technique selected and reason for selecting the techniques
 - Additional Data Processing:

Analyzing data is a non-linear process, so additional data cleaning and preparation may need to be performed on the cleaned data from Proposal.

Any additional data processing should be described and relevant R code shared.
4. Evaluate and scrutinize the results.
 - How well can you answer your questions? How accurate are the results on the scale of the problem?

Project: Final Report – Some Steps to Consider

5. Revise your analyses

- What should be changed to reflect your initial findings? What more is needed to better understand the problem?

6. Report your conclusions

- Draft a document that places the problem in the proper context. Describe your analyses in language and technical details that are appropriate for your intended audience. Decide what kind of results to include.

7. Recommendations

- Place your conclusions in the context of the domain or application.
- What should be done differently? What specific choices should decision-makers consider? What kinds of improvements are likely, and how can you quantify them?
- Keep in mind that not every observed difference (e.g. demographic differences) is directly actionable. As much as possible, frame your analysis to reflect choices that are within the control of your intended audience.

- The steps outlined above are a recommended process, not a recipe.

Your final report should focus on the questions, analyses, and recommendations that you settled on without necessarily conveying all of the work of the earlier iterations. Keep in mind that certain results that may be helpful for you in understanding the data may be less relevant to decision-makers. Try to create a report that provides a clear and thorough analysis

Project: Final Report – Assessment (Total 200 points)

<u>Criteria</u>	<u>Points</u>
Selected an appropriate analytical technique	40
Demonstrated knowledge of R tools to implement the analytical techniques used	40
Properly interpreted results from analysis	40
Conclusions for analysis and recommendations address the question	40
Write-up is concise and persuasive, and suitable for an audience of non-technical decision makers	40

Project: Final Report - Submission

What to submit - submit the following as part of the final report:

1. Data: Submit the raw data and cleaned data turned in with the Proposal. If the data is too large, post the data to a cloud drive and share a persistent link. You should not expect the teaching staff to create accounts or log in to websites to download the data. If you modified the cleaned data, then also share the processed data and R code used to obtain the processed data from the cleaned data.
2. Working R code for Analysis: Share R code used to run the various analysis attempted. This block of R code should contain all the different analyses conducted, those that were used for the final conclusions and the ones that were not. The R code may be submitted as an R script (.R), R Markdown (.rmd), or a knit R Markdown (.html) file.
3. Final R Code: Share R code that contains only the final analysis. In addition, this should also include code for any charts created to corroborate the conclusions. The R code may be submitted as an R script (.R), R Markdown (.rmd), or a knit R Markdown (.html) file.

Project: Final Report - Submission

4. Report (text, Word, PDF, or HTML format) (5-10 written pages, no limits on spacing and font size):

The report should include

- Statement of the research problem or question(s) being addressed
- Briefly discuss the data and justify its suitability
- Reasons behind the choice of analytical technique(s). Explain why the technique is suitable for the question(s) being addressed.
- Discuss the results from the analyses run. Wherever possible use charts to convey results succinctly.
- Discuss the conclusions from the analysis and offer recommendations in a form that is simple and understandable to decision makers. Support conclusions with relevant charts

Project: Final Report– Due April 19st at 11:59 pm

How to submit:

- Only one member of your group need submit this assignment on behalf of the group.
- To complete your submission,
 - i. Click the blue Submit Assignment button at the top of the Project: Final Report page under module 14.
 - ii. Click the Choose File button, and locate your submission. Repeat for each file you submit.
 - iii. Feel free to include a comment with your submission.
 - iv. Finally, click the blue Submit Assignment button.

(depending on canvas web design update, this process may change)

Project Presentation – Overview

- At the end of the course, each group will design and deliver a brief presentation that explains analytical outputs to a general audience.
- In class on April 19th
- Length of the presentation (5 -7 mins, depending on the number of groups)

Project Presentation - Assessment

<u>Criteria</u>	<u>Points</u>
Design an impactful presentation	20
Deliver and explain analytical outputs to a general audience	15
Maintain a constructive presence when debating data and analytical results	15

Project Presentation – Submission (Due April 18th at 12:00 pm [at noon])

- What to Submit:

- A slide deck (in Powerpoint)

- How to Submit:

- Only one member of your group need submit this assignment on behalf of the group.
 - To complete your submission:
 - i. Click the blue **Submit Assignment** button at the top of the Project Presentation page under module 14.
 - ii. Click the **Choose File** button, and locate your submission. Repeat for each file you submit.
 - iii. Feel free to include a comment with your submission.
 - iv. Finally, click the blue **Submit Assignment** button.
- (depending on canvas web design update, this process may change)

Getting Ready

Getting Ready

- Review Framework and Methods – 1 material
- Check version: R and R Studio
- Signing up for group projects

1. Working with R

Read in a data.frame

Text files with comma as a delimiter may be read using `read.csv()`. To read in a file, one must use one of the two methods:

- Method 1: Specify the full path of the file in `read.csv()`, or
- Method 2: Set the working directory to the location of the file. Then specify file name in `read.csv()`

To find the file path:

- For Windows: With Shift pressed, right click on file. Select "Copy as Path"
- For Mac: Right click on file and press Option. Select "Copy xxx as Pathname"

Here is an illustration on how to read in a file called `info.csv` that is saved in the folder,

`c:/my_classes/best_course_ever` (or `/my_classes/best_course_ever` on a Mac) using `read.csv()`

Method 1

```
read.csv('c:/my_classes/best_course_ever/info.csv') # on Windows  
read.csv('/my_classes/best_course_ever/info.csv') # on Mac
```

Method 2

```
setwd('c:/my_classes/best_course_ever/') # on Windows  
setwd('/my_classes/best_course_ever/') # on Mac  
  
read.csv('info.csv')
```

Packages

The R System involves a Base System and a set of Packages. A Package is a collection of code and functions written for the R language. Usually focuses on a specific task or problem. Implementation is described in the package documentation available on [CRAN](#). By only installing packages that one needs, R can be kept light. Most useful R applications appear in packages.

To use a package, install it only once. Then call the package for each R session. Here is an illustration with the [ggplot2](#) package.

```
install.packages('ggplot2')
```

To load into current R session

```
library(ggplot2)
```

As noted above, a package needs to only be installed once. To use the package call it using [library\(\)](#) in each R session.

Versions

The software R and its libraries are constantly updated in order to add new capabilities and fix bugs. Such changes can affect the results of functions. The code in this class was generated using R version 4.1.2 (2021-11-01) and the active version of libraries as of Dec 22, 2021. To ensure you are able to generate the same output as in the R code for class, please ensure you are using R version 4.1.2 (2021-11-01).

Check Version : R

Session: New Session

The screenshot shows the RStudio interface with the 'Session' menu highlighted by a blue arrow. The terminal window displays the R version information:

```
R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

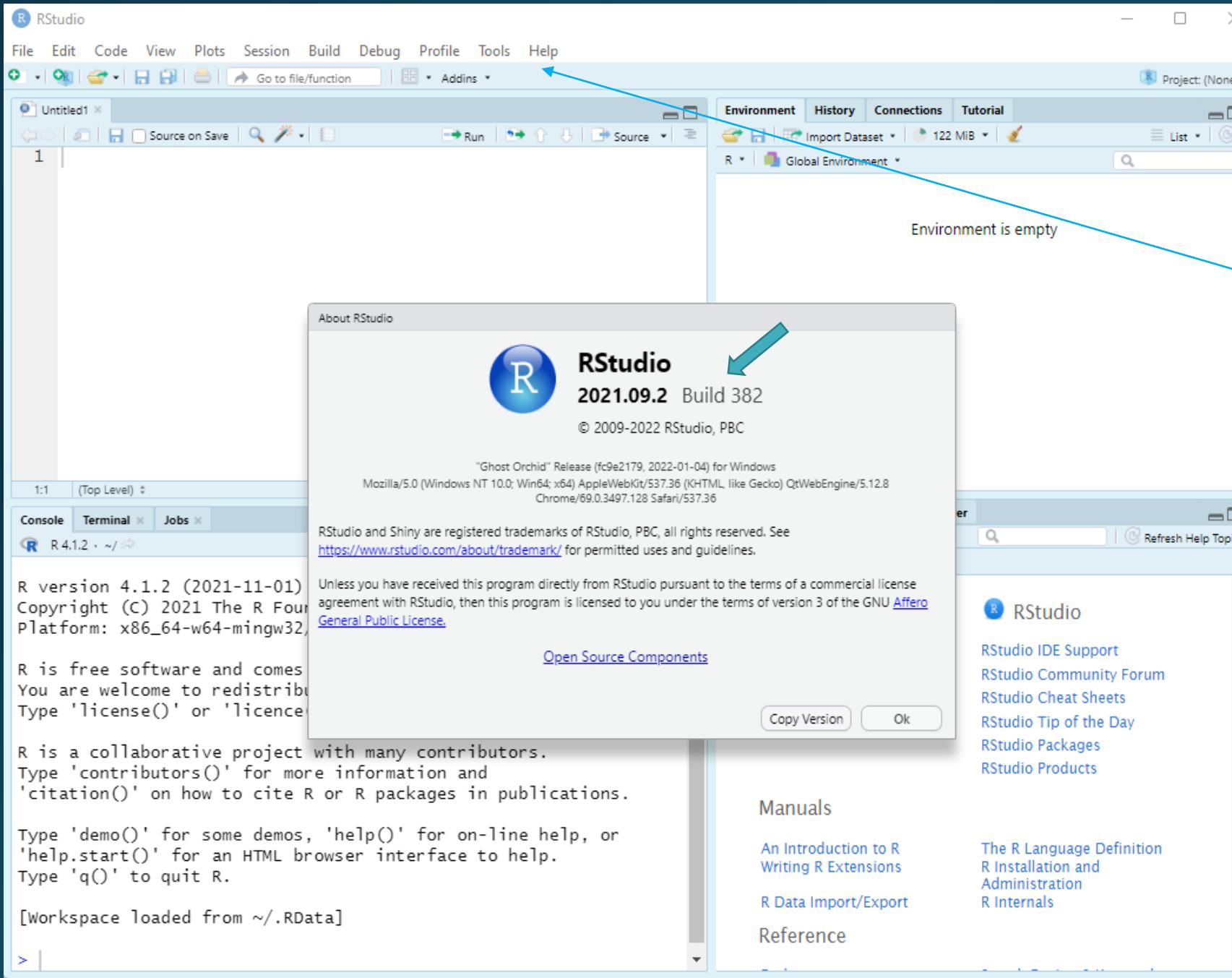
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]
```

The screenshot shows the R console window with the command `> version` highlighted by a blue arrow. The output shows the R version and its configuration:

```
platform      x86_64-w64-mingw32
arch         x86_64
os          mingw32
system      x86_64, mingw32
status
major        4
minor        1.2
year        2021
month       11
day         01
svn rev    81115
language     R
version.string R version 4.1.2 (2021-11-01)
nickname   Bird Hippie
```



Check Version : RStudio

For Window User
Help: About RStudio

For Mac User
RStudio: About RStudio

Setup

- Ensure you have
 - R (version 4.1.2, (2021-11-01, Bird Hippie)) and
 - RStudio (version 2021.09.2+382)
- Latest version of R:
 - <https://cran.r-project.org/>
- Latest version of Rstudio Desktop
 - <https://www.rstudio.com/products/rstudio/download/>

← → C cran.r-project.org

Apps CU Admin General Daily Depo

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2021-11-01, Bird Hippie) [R-4.1.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

- **Install R**
- <https://cran.r-project.org/>

rstudio.com/products/rstudio/download/

Apps CU Admin General Daily Depo

R Studio

Products Solutions Customers Resources About Pricing

Download the RStudio IDE

Choose Your Version

The RStudio IDE is a set of integrated tools designed to help you be more productive with R and Python. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace.

[LEARN MORE ABOUT THE RSTUDIO IDE](#)

RStudio Desktop **RStudio Desktop Pro** **RStudio Server** **RStudio Workbench**

RStudio Desktop	RStudio Desktop Pro	RStudio Server	RStudio Workbench
Open Source License	Commercial License	Open Source License	Commercial License
Free	\$995	Free	\$4,975
/year		/year (5 Named Users)	
DOWNLOAD	BUY	DOWNLOAD	BUY
Learn more	Learn more	Learn more	Evaluation Learn more
Integrated Tools for R	✓	✓	✓
Priority Support	✓		✓
Access via Web Browser		✓	✓
RStudio Professional Drivers	✓		✓

R Studio Team
RStudio's recommended professional data science solution for every team. RStudio Team is a bundle of RStudio's popular professional software for data analysis, package management, and sharing data products.
[Learn more about RStudio Team](#)

▪ Install RStudio Desktop

<https://www.rstudio.com/products/rstudio/download/>

RStudio Desktop 2021.09.2+382 [Release Notes](#)

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:

 [DOWNLOAD RSTUDIO FOR WINDOWS](#)
2021.09.2+382 | 156.89MB

Requires Windows 10 (64-bit)



All Installers

Linux users may need to import RStudio's public code-signing key [prior to installation](#), depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an older version of RStudio.

OS	Download	Size	SHA-256
Windows 10	 RStudio-2021.09.2-382.exe	156.89 MB	7f957be8
macOS 10.14+	 RStudio-2021.09.2-382.dmg	204.09 MB	ae18a925
Ubuntu 18/Debian 10	 rstudio-2021.09.2-382-amd64.deb	117.15 MB	f3dd8823
Fedora 19/Red Hat 7	 rstudio-2021.09.2-382-x86_64.rpm	133.82 MB	a1190f21
Fedora 28/Red Hat 8	 rstudio-2021.09.2-382-x86_64.rpm	133.85 MB	ce832b22
Debian 9	 rstudio-2021.09.2-382-amd64.deb	117.43 MB	2c9c4a33
OpenSUSE 15	 rstudio-2021.09.2-382-x86_64.rpm	118.68 MB	6e2b6848

In instances when updating from an older version to a newer version, sometimes elements of the older version may be retained. One such issue is with the Random Number Generator used for creating pseudo-random numbers. To verify that your computer is using the latest random number generator, your output from running `RNGkind()` should be as follows.

```
RNGkind()
```

```
## [1] "Mersenne-Twister" "Inversion"      "Rejection"
```

If you got the above output, then you can stop here, your random number generator is the latest.

When updating from an older version of R to the latest (version 4.0. as of October 12, 2020), sometimes the random number generator does not update to the latest version. If this is the case for you, then the reason for this is that RStudio is using the random number generator from a saved .RData file. To address this issue, save all your unsaved files in RStudio, close RStudio. Restart RStudio and run the following script to remove your .rdata file in your working directory.

```
file.remove('.RData')
q('no')
```

When you are done, the output from running RNGkind() should look as below

```
RNGkind()
## [1] "Mersenne-Twister" "Inversion"      "Rejection"
```

If you got the above output, you don't have to repeat this exercise. Steps above only need to be done once.

Getting Help

The Universe of R is vast and expanding rapidly as new packages get added to CRAN. As a result, it is humanly impossible to learn and remember how to use each function.

To get help, you could

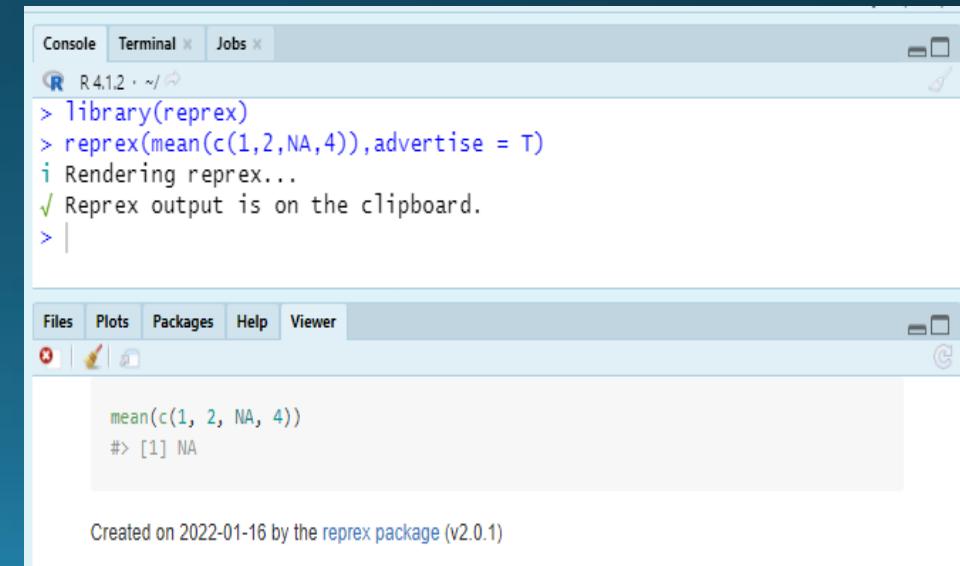
- Search for answers
 - Access R Help by typing "?" before a function (e.g., ?mean). See Examples
 - Keyword search on [Stack Overflow](#), [R Bloggers](#), or R Mailing List: r-help@r-project.org
 - But, by far the best is Google: (a) copy-paste error messages from R, or (b) keyword search for function followed by R
- Ask for Help
 - Class Discussions Board, Stack Overflow, Github
 - But, be sure to share a reproducible example

When asking for help, clearly state the problem and most importantly, provide a reproducible example of the problem. This may require using a toy dataset (e.g., iris, or mtcars) to illustrate the problem. The [reprex library](#) offers a handy function that generates code and result in the clipboard to make it easy to paste into Discussions or Stack Overflow. To illustrate, running the function below will generate an output that is ready for pasting into a forum.

```
library(reprex)
reprex(mean(c(1,2,NA,4)),advertise = F)
```

Here is the paste-ready result.

```
mean(c(1, 2, NA, 4))
#> [1] NA
```



Signing up for Project Groups

Signing Up for Project Groups

#	Project Topic	Owner (Topic Initiator)	Member 1 - Full Name	Member 1 - UNI	Member 2 - Full Name	Member 2 - UNI	Member 3 - Full Name	Member 3 - UNI	Member 4 - Full Name	Member 4 - UNI	Member 5 - Full
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16	Undecided										
17											
18											
19											
20											
21											

- Propose a topic with a complete team and enter the information in the spreadsheet
- Propose a topic and look for additional team members using the spreadsheet
- Contact the topic “owner” to discuss joining the team
- Write to us to help you to join a team