

CT Academy | Reading and Viewing List

Data Insight Fundamentals

Warm Up

- **Article:** These AI-invented paint color names are so bad they're good
 - <https://www.curbed.com/2017/5/31/15700260/paint-colors-sherwin-williams-benjamin-moore>
- **Article:** Math Model Determines Who Wrote Beatles' "In My Life": Lennon or McCartney
 - <https://www.inverse.com/article/47568-beatles-song-statistics-lennon-mccartney>
- **Article:** Amazon scraps secret AI recruiting tool that showed bias against women
 - <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MKo8G>
- **Video:** The Human Insights Missing from Big Data
 - <https://www.youtube.com/watch?v=pk35J2u8KqY>

Getting Going

- **Video:** Data Attributes (Part 1) | Introduction to Data Mining part 2
 - <https://www.youtube.com/watch?v=hu7iGGnzg3Y>
- **Video:** Data Attributes (part 2) | Introduction to Data Mining part 3
 - https://www.youtube.com/watch?v=xrFtN_UJhYc
- **Video:** Pi Analytics example: Would you eat the mushroom?
 - <https://www.youtube.com/watch?v=7mPJqw78oWQ>
- **Video:** Machine Learning Basics | What Is Machine Learning? | Introduction To Machine Learning
 - <https://www.youtube.com/watch?v=ukzFl9rgwfU&app=desktop>
- **Article:** The AI Hierarchy of Needs
 - <https://hackernoon.com/the-ai-hierarchy-of-needs-18f11fcc007>
- **Video:** The Analytics Translator
 - <https://www.youtube.com/watch?v=rPwjFtOZObo>
- **Video:** The Must-Have Future Career: Analytics Translator
 - <https://www.youtube.com/watch?v=6lHq-2HNUvI>

Final Push

- **Video:** The era of blind faith in big data must end | Cathy O'Neil
 - https://www.youtube.com/watch?v=_2u_eHHzRto
- **Article:** "But its already public, right?": The ethics of using online data
 - PDF uploaded to Moodle Site
- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 2 (Fundamentals of Data Insight, chapters 13 and 14)
 - <https://idlewyldanalytics.com/vol2>

Data Collection and Processing

Warm Up

- **Video:** Sampling Design Process
 - <https://www.youtube.com/watch?v=4e5VH9NKNVY>
- **Video:** 4.2 Probability Sampling Techniques
 - <https://www.youtube.com/watch?v=jauhoR7w1YM>
- **Video:** REST API concepts and examples
 - <https://www.youtube.com/watch?v=7YcW25PHnAA>

Getting Going

- **Article:** Quantitative Data Cleaning for Large Databases
 - <http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf>
- **Article:** Data Quality Strategy: a Step-by-Step approach
 - <https://archive.sap.com/kmuuid2/b0b02c2a-6364-2b10-8b9b-e29a960f411e/Data%20Quality%20Strategy%3A%20A%20Step-by-Step%20Approach.pdf>
- **Video:** The Trouble with Missing Data - Computerphile
 - <https://www.youtube.com/watch?v=oCQbC818KKU>
- **Video:** Missing Data Mechanisms
 - <https://www.youtube.com/watch?v=XnnA9z7lv4Q>
- **Video:** Dealing With Missing Data - Multiple Imputation
 - <https://www.youtube.com/watch?v=LMsULWGtP2c>

Final Push

- **Video:** The Curse of Dimensionality
 - <https://www.youtube.com/watch?v=QZoDtNFdDko>
- **Video:** Curse of Dimensionality Two - Georgia Tech - Machine Learning
 - <https://www.youtube.com/watch?v=OyPcbeiwps8>
- **Video:** Data Analysis 5: Data Reduction - Computerphile
 - <https://www.youtube.com/watch?v=8k56bvhXw4s>
- **Video:** Principal Component Analysis
 - <https://www.youtube.com/watch?v=TJdH6rPA-TI>

Bonus Material and Additional References

- **Video:** How to Build A Spider With Python (Web Scraping)
 - <https://www.youtube.com/watch?v=qj6H38Xwvu0>
- **Video:** Introduction to Regular Expressions
 - <https://www.youtube.com/watch?v=vcRPNhLbhoc>
- **Article:** Missing Data
 - https://en.wikipedia.org/wiki/Missing_data
- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 2 (Fundamentals of Data Insight, chapters 15 and 16)
 - <https://idlewyldanalytics.com/vol2>

Data Exploration and Visualization

Warm Up

- **Video:** The Art of Data Visualization | Off Book | PBS Digital Studios
 - <https://www.youtube.com/watch?v=AdSZJzb-aX8>
- **Video:** The beauty of data visualization - David McCandless
 - <https://www.youtube.com/watch?v=5Zq-C8AAlGg>

Getting Going

- **Article:** One Dataset, Visualized 25 Ways
 - <https://flowingdata.com/2017/01/24/one-dataset-visualized-25-ways/>
- **Video:** Data Analysis 2: Data Visualisation - Computerphile
 - <https://www.youtube.com/watch?v=j6CIFJhnG18>
- **Video:** The best stats you've ever seen
 - https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen
- **Video:** How a blind astronomer found a way to hear the stars | Wanda Diaz Merced
 - <https://www.youtube.com/watch?v=-hY9QSdaReY>
- **Article:** Turning hurricanes into music
 - <https://earthsky.org/earth/turning-hurricanes-storms-into-music-sonification>
- **Video:** Sonification of Storms Year 2005 - Mark Ballora
 - <https://www.youtube.com/watch?v=TKTLE1rRUDA&feature=youtu.be>

Final Push

- **Article:** Perception in Visualization
 - <https://www.csc2.ncsu.edu/faculty/healey/PP/>
- **Article:** Ask the question, Visualize the answer
 - <https://flowingdata.com/2018/10/17/ask-the-question-visualize-the-answer/>
- **Article:** The death of interactive infographics?
 - <https://medium.com/@dominikus/the-end-of-interactive-visualizations-52c585dcafcfb#.ya4dn1mzh>
- **Article:** Error bars in bar charts. You probably shouldn't
 - <https://gorelik.net/2019/10/07/error-bars-in-bar-charts-you-probably-shouldnt/>
- **Article:** 50 Great Examples of Data Visualization
 - <https://www.webdesignerdepot.com/2009/06/50-great-examples-of-data-visualization/>

Bonus and Reference

- **Video:** How a Fitness App's Heat Map Uncovers Military Bases | NYT
 - <https://www.youtube.com/watch?v=0lB8p-YpXwA>
- **Article:** Our solar system sounds like a depressed Radiohead cover band
 - <https://theoutline.com/post/3573/radiohead-true-love-waits-solar-system?zd=1&zi=uzp7waba>
- **Article:** Data Visualization: Visualization Types
 - https://guides.library.duke.edu/datavis/vis_types
- **Textbook:** The Practice of Data Visualization
 - <https://practicedataviz.com>

Statistical and Mathematical Foundations

Warm Up

- **Video:** Why you should love statistics | Alan Smith
 - <https://www.youtube.com/watch?v=ogeGJS0GEF4>
- **Video:** What Is Statistics: Crash Course Statistics #1
 - <https://www.youtube.com/watch?v=sxQaBpKfDRk>
- **Video:** Statistics: The average | Descriptive statistics | Probability and Statistics | Khan Academy
 - https://www.youtube.com/watch?v=uhxtUt_-GyM&list=PL87D6C3431177ED5C
- **Video:** Vectors, what even are they? | Essence of linear algebra, chapter 1
 - https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFi_tgF8hE_ab
- **Article:** Basic Matrix Operations
 - https://www.georgebrown.ca/uploadedFiles/TLC/_documents/Basic%20Matrix%20Operations.pdf

Getting Going

- **Article:** Statistical Distributions
 - http://pages.stern.nyu.edu/~adamodar/New_Home_Page/StatFile/statdistns.htm
- **Video:** Central limit theorem
 - <https://www.youtube.com/watch?v=b5xQmk9veZ4>
- **Video:** The Applications of Matrices MajorPrep
 - <https://www.youtube.com/watch?v=rowWM-MijXU> 25:15
- **Article:** Linear Algebra and Matrix Decompositions
 - <https://people.duke.edu/~ccc14/sta-663/LinearAlgebraMatrixDecompWithSolutions.html>

Final Push

- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 1 (Prelude to Data Understanding)
 - <https://idlewyldanalytics.com/vol1>
- **Article:** Linear Regression Example
 - <https://stattrek.com/regression/regression-example.aspx?Tutorial=AP>
- **Video:** 13.1 The general linear model
 - <https://www.youtube.com/watch?v=PY-xznrFNPI>
- **Article:** Generalized Linear Models
 - <https://web.as.uky.edu/statistics/users/pbreheny/760/S13/notes/1-24.pdf>

Bonus Video: What's a Tensor?

- <https://www.youtube.com/watch?v=f5liqUkoZTw>

Data Information Architectures

Warm Up

- **Video:** What is Big Data|Big Data Types|Types of Data|Structured Data|Unstructured Data
 - Start at 3:06: <https://www.youtube.com/watch?v=dK4aGzeBPkk>
- **Video:** Database Tutorial for Beginners
 - <https://www.youtube.com/watch?v=wRojgoeOsZA>
- **Video:** Entity Relationship Diagram (ERD) Tutorial - Part 1
 - <https://www.youtube.com/watch?v=QpdhBUYk7Kk>
- **Video:** MySQL Tutorial for Beginners [2019] - Full Course
 - Up to 4:20: https://www.youtube.com/watch?v=7S_tz1z_5bA
- **Video:** What is Metadata?
 - <https://www.youtube.com/watch?v=HXAstVP3-yo>
- **Video:** Database VS Data Warehouse
 - <https://www.youtube.com/watch?v=FxQG65OhXAO>

Getting Going

- **Video:** What is a Data Model?
 - <https://www.youtube.com/watch?v=4qFZ-5i4GS8>
- **Video:** What is NoSQL?
 - https://www.youtube.com/watch?v=BgQFJ_UNIgw
- **Video:** Data Lake VS Data Warehouse
 - <https://www.youtube.com/watch?v=AwbKwcv7bqg>
- **Video:** MongoDB in 5 Minutes with Eliot Horowitz
 - <https://www.youtube.com/watch?v=EE8ZTQxaoAM>
- **Video:** Graph Databases 101 | DataStax
 - <https://www.youtube.com/watch?v=vJcxRjJ982k>
- **Video:** What is Hadoop?: SQL Comparison
 - <https://www.youtube.com/watch?v=MfF75oYVDxM>

Final Push

- **Video:** Tim Berners-Lee: The next Web of open, linked data
 - https://www.youtube.com/watch?v=OM6XIIcm_qo
- **Video:** Intro to the Semantic Web
 - <https://www.youtube.com/watch?v=OGg8A2zfWKg>
- **Video:** An Introduction to the Semantic Web
 - <https://www.youtube.com/watch?v=V6BR9DrmUOA>
- **Video:** SPARQL in 11 minutes
 - <https://www.youtube.com/watch?v=FvGndkpa4Ko>
- **Article:** Ontology Development 101: A Guide to Creating Your First Ontology
 - https://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- **Video:** RDFa Basics
 - <https://www.youtube.com/watch?v=zteyEk9LADs>

Bonus

- **Video:** W3School's SQL Tutorial
 - <https://www.w3schools.com/sql/default.asp>

Basic Data Analysis

Warm Up

- **Video:** What is Business Analytics?
 - <https://www.youtube.com/watch?v=7R2ebocENms>
- **Video:** What is Descriptive Analytics? - Data Science Wednesday – YouTube
 - https://www.youtube.com/watch?v=cHGhSj_Ax-o
- **Video:** Data Analysis 1: What is Data? - Computerphile – YouTube
 - <https://www.youtube.com/watch?v=SEeQgNdJ6AQ>
- **Video:** Statistics 101: Describing a Categorical Variable – YouTube
 - <https://www.youtube.com/watch?v=vrWYw8d283o>
- **Video:** INTERACTIVE: Part 1: Identify the Independent and Dependent Variables with the MythBusters! – YouTube
 - <https://www.youtube.com/watch?v=lojTMDtX4WY>

Getting Going

- **Article:** setwd / getwd – Setting The Working Directory In R
 - <http://www.programmingr.com/examples/neat-tricks/setwd-getwd-setting-the-working-directory-in-r/>
- **Article:** This R Data Import Tutorial Is Everything You Need (article) – DataCamp
 - <https://www.datacamp.com/community/tutorials/r-data-import-tutorial>
- **Article:** Beginner's guide to R: Easy ways to do basic data analysis | Computerworld
 - <https://www.computerworld.com/article/2598083/app-development-beginner-s-guide-to-r-easy-ways-to-do-basic-data-analysis.html>
- **Article:** Exploratory Data Analysis in R for beginners (Part 1)
 - <https://towardsdatascience.com/exploratory-data-analysis-in-r-for-beginners-fe031add7072>
- **Article:** Frequencies and CrossTabs
 - <https://www.statmethods.net/stats/frequencies.html>

Final Push

- **Article:** How to properly tell a story with data – and common pitfalls to avoid.
 - <https://towardsdatascience.com/how-to-properly-tell-a-story-with-data-and-common-pitfalls-to-avoid-317d8817e0c9>
- **Website:** Data visualization with base R
 - <https://www.r-graph-gallery.com/base-R.html>
- Note: We will be going into more depth on how to use ggplot2 in week 6, but you may choose to get a start here: **Article:** R for Journalists: Chapter 4 - Visualizing Data
 - <https://learn.r-journalism.com/en/visualizing/>

Machine Learning

Warm Up

- **Article:** Introduction to Machine Learning
 - <https://www.idlewyldanalytics.com/chapters/DUDADS-ch19.pdf>

Getting Going

- **Video:** Machine Learning – Supervised VS Unsupervised Learning
 - <https://www.youtube.com/watch?v=cfj6yaYE86U>
- **Vidéo:** APPRENTISSAGE SUPERVISÉ : LES 4 ÉTAPES - ML#2
 - <https://www.youtube.com/watch?v=K9zoOD22My4>
- **Vidéo:** Apprentissage automatique [1.3] : Concepts fondamentaux - types d'apprentissage
 - <https://www.youtube.com/watch?v=4EvNjHyepRY>
- **Article:** Introduction to semi-supervised learning and adversarial training
 - <https://medium.com/inside-machine-learning/placeholder-3557ebb3d470>
- **Article:** Complete guide to Association Rules (1/2)
 - <https://towardsdatascience.com/association-rules-2-aa9a77241654>
- **Article:** Complete guide to Association Rules (2/2)
 - <https://towardsdatascience.com/complete-guide-to-association-rules-2-2-c92072b56c84>
- **Video:** Learning to See, parts 1 to 15 (NOTE: they say that when the Beatles released *Sgt. Pepper's Lonely Hearts Club Band*, contemporary music groups thought that it was pointless to continue to work on their records. On an unrelated note, take the time to go through this astounding series of videos by Welch Labs. It's a bit long, but it is completely worth it.)
 - <https://www.youtube.com/playlist?list=PLiaHhY2iBX9ihLasvE8BKnS2Xg8AhY6iV>
- **Article:** The 5 Clustering Algorithms Data Scientists Need to Know
 - <https://www.towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>
- **Video:** Unsupervised Learning Explained
 - <https://www.youtube.com/watch?v=IEfrroYr684>
- **Video:** Unsupervised Machine Learning: Crash Course Statistics #37
 - <https://www.youtube.com/watch?v=UN8k5zSl6g>

Reference Works

- <https://faculty.marshall.usc.edu/gareth-james/ISL> (G. James, D. Witten, T. Hastie, R. Tibshirani)
- <https://github.com/asadoughi/stat-learning> (unofficial solution set, A. Sadoughi)
- <https://web.stanford.edu/~hastie/ElemStatLearn/> (T. Hastie, R. Tibshirani, J. Friedman)
- https://waxworksmath.com/Authors/G_M/Hastie/WriteUp/Weatherwax_Epstein_Hastie_Solution_Manual.pdf (unofficial solution manual, J.L. Weatherwax, D. Epstein)
- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 3 (Spotlight on Machine Learning, <https://idlewyldanalytics.com/vol3>)

Predictive Analytics

Warm Up

- **Video:** The Fundamentals of Predictive Analytics - Data Science Wednesday
 - <https://www.youtube.com/watch?v=4y6fUC56KPw>
- **Video:** A Brief History of Predictive Analytics – Parts 1, 2 and 3
 - <http://afterinc.com/brief-history-predictive-analytics-part-1/>
 - <http://afterinc.com/brief-history-predictive-analytics-part-2/>
 - <http://afterinc.com/brief-history-predictive-analytics-part-3/>

Getting Going

- **Video:** The Discovery of Statistical Regression:
 - <https://priceconomics.com/the-discovery-of-statistical-regression/>
- **Video:** Regression I: What is regression? | SSE, SSR, SST | R-squared | Errors (ϵ vs. e)
 - <https://www.youtube.com/watch?v=aq8VU5KLmkY>
- **Video:** Time Series Forecasting in Minutes
 - https://www.youtube.com/watch?v=wGUV_XqchbE
- **Video:** Time Series Prediction
 - https://www.youtube.com/watch?v=d4Sn6ny_5LI

Final Push

- Neural Networks Demystified, Parts 1 - 7
 - <https://www.youtube.com/watch?v=bx2T-V8XRrs>
 - <https://www.youtube.com/watch?v=UJwK6jAStmg>
 - <https://www.youtube.com/watch?v=5u0jaA3qAGk>
 - <https://www.youtube.com/watch?v=GlcnxUlrtk>
 - <https://www.youtube.com/watch?v=pHMzNW8Agq4>
 - <https://www.youtube.com/watch?v=gKM9Td6RVgQ>
 - <https://www.youtube.com/watch?v=S4ZUwgesjS8>

Bonus

- **Video:** What is Multicollinearity? Extensive video + simulation!
 - <https://www.youtube.com/watch?v=Cba9LJ9IS8s>
- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 4 (Techniques of Data Analysis, chapter 25)
 - <https://idlewyldanalytics.com/vol4>
- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 4 (Prelude to Data Understanding, chapters 8-9)
 - <https://idlewyldanalytics.com/vol1>

Text Mining and Sentiment Analysis

Getting Going

- **Video:** Natural Language Processing: Crash Course Computer Science #36
 - <https://www.youtube.com/watch?v=fOvTtapxa9c>
- **Tutorial:** Sentiment Analysis in 4 Minutes
 - <https://www.youtube.com/watch?v=AJVP96tAWxw>
- **Video:** Intro to Text Mining – Interview with Rada Mihalcea
 - <https://www.youtube.com/watch?v=cjouCoj4lp8>

Final Push

- **Tutorial:** Data Science 101: Sentiment Analysis in R Tutorial
 - <http://blog.kaggle.com/2017/10/05/data-science-101-sentiment-analysis-in-r-tutorial/>
- **Tutorial:** Text Mining in Python: Steps and Examples
 - <https://medium.com/towards-artificial-intelligence/text-mining-in-python-steps-and-examples-78b3f8fd913b>
- **Video:** Katharine Jarmul – I Hate You, NLP... ;)
 - <https://www.youtube.com/watch?v=vitEXiOuiEk>
- **Video:** LDA Topic Models
 - <https://www.youtube.com/watch?v=3mHy4OSyRfo>
- **Video:** How to Build a Text Mining, Machine Learning Document Classification System in R!
 - <https://www.youtube.com/watch?v=j1V2McKbkLo&list=PLRQrw2NjzkqPSBNCCXbnCHgUVrr35cm6B&index=3>

Reference Works

- <https://www.tidyttextmining.com/> (J. Silge, D. Robinson)
- <https://github.com/dgrtwo/tidy-text-mining>
- <https://www.nltk.org/book/> (S. Bird, E. Klein, E. Loper)
- <https://groups.google.com/forum/#!forum/nltk-users>
- **Textbook:** Data Understanding, Data Analysis, and Data Science, volume 4 (Techniques of Data Analysis, chapter 27)
 - <https://idlewylanalytics.com/vol4>

Decision-Making and Evaluating Outcomes

Specific Resources

- A Framework for Ethical Decision-Making
 - <https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/>
- 3 Frameworks for Ethical Decision-Making
 - <https://www.floridatechonline.com/blog/business/3-frameworks-for-ethical-decision-making/>
- Outcome Management Guide and Tools
 - <https://www.canada.ca/en/treasury-board-secretariat/services/information-technology-project-management/project-management/outcome-management-guide-tools.html>
- GoC Policy on Results
 - <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=31300>

Reference Works

- How the Dot Collector Works (R. Dallio)
 - <https://www.youtube.com/watch?v=tUu-6UxuGr8>
- How to Decide: Simple Tools for Making Better Choices (A. Duke)
- The Signal and the Noise (N. Silver)
- Superforecasting: The Art and Science of Prediction, (P.E. Tetlock)

Government of Canada Legislation, Policies and Directives

Policy drivers for ethical handling of data

- Federal Data Strategy Roadmap / 2023–2026 Data Strategy for the Federal Public Service
 - <https://www.canada.ca/en/treasury-board-secretariat/corporate/reports/2023-2026-data-strategy.html>
- Canada's Digital Ambition 2022
 - <https://www.canada.ca/en/government/system/digital-government/government-canada-digital-operations-strategic-plans/canada-digital-ambition.html>

Policy drivers for ethical handling of Indigenous data

- UNDRIP / UNDA / UNDA Action Plan
 - <https://www.justice.gc.ca/eng/declaration/index.html>
 - <https://afn.ca/rights-justice/united-nations-declaration-on-the-rights-of-indigenous-peoples/>
- Departmental Reconciliation Strategy
 - DFO: <https://www.dfo-mpo.gc.ca/fisheries-peches/aboriginal-autochtones/reconciliation-eng.html>
 - CIRNAC: <https://www.rcaanc-cirnac.gc.ca/eng/1675097236893/1675097328140>

Policy drivers for responsible use of A.I.

- Responsible use of artificial intelligence
 - <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html>
- TBS Directive on Automated Decision-Making
 - <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>
- Government of Canada Digital Standards: Playbook
 - <https://www.canada.ca/en/government/system/digital-government/government-canada-digital-standards.html>