Data management and visualization

Week 1

Tile 1: A data analytics pipeline

- **ppt**

  -Generic Data Analytics Pipeline

  https://docs.google.com/presentation/d/19-bJFLGgen-Lg0-n9qp4LHSNvkPJe2eYly1mw1_j6zM/edit

- **Docx**

  -research data management lifecycle

  -US Geological Survey Data Lifecycle

  -Enterprise Data Management

  -model for Data Maturity

  -CRISP-DM (data mining)

  https://docs.google.com/presentation/d/19-bJFLGgen-Lg0-n9qp4LHSNvkPJe2eYly1mw1_j6zM/edit#slide=id.p

Week 2

Tile 2: What is data?

- **ppt**

  -Data slide 3

  -Structured vs Unstructured 4
  -Quantitative vs Qualitative 6
  -Discrete vs Continuous 7
  -Four levels of data 8

  1. Nominal
  2. Ordinal
  3. Interval
  4. Ratio

- Special types of data 5

- Data sources 12

- Open data 19

https://docs.google.com/presentation/d/1vC8iNCdWCHt_ugn0Mb9KWe4epKmzrYIPpHISJUOvisY/edit#slide=id.ga17f6894f7_0_312

- **Docx**

- Files

1. Text or binary.
2. Open or proprietary.
3. Structured or unstructured.
   - Spreadsheets & File Types
   - Open Data
   - Linked Data
   - Uniform resource identifier
   - SPARQL Protocol and RDF Query Language

https://docs.google.com/document/d/1W31RGUGYEi0Ha0xwxqWUDA15jG-BgAlPhwf4zJWR0hI/edit#heading=h.1d2986nrw3pc

Week 3

Tile 3: Describing data?

- **ppt**

- Metadata 12

- Types of metadata 15
- Abstraction hierarchy 21

- Granularity 22

- Where does metadata come from? 24

https://docs.google.com/presentation/d/13nKtTmEX52tTJQbgb6nZ7YBcWb-UW_ow5TuHgsAR9DY/edit#slide=id.g15b3823fa0f_0_624

- **Docx**

- Metadata
- What is it and why is it useful?
- Metadata Granularity
- Metadata Standards

- How is metadata created?
https://docs.google.com/document/d/1JF7sBg0Es2NYaqg3_jVPt5cEYKH_qiWsxjATxjbQkrI/edit#heading=h.vany23ko44a

- **ppt**

Overview of Big Data (3Vs)

- Big data 3
- Big data: Volume 4
- Big data: Variety 7
- Big data: Velocity 8
- Big data: Veracity 9

https://drive.google.com/file/d/1jIN2cF2453Bsi_Dm0kzflQuv_O2wAAOZ/view

- **Docx**

- Big Data
- The Internet Minute
- Instrumented Vehicles
- The NY Stock Exchange

https://docs.google.com/document/d/1XoKCaZUqsAu1D_vANsAmsDCZTioyKCnj9rcj4fPEUJA/edit#heading=h.4wtucephf0nw

Week 5

Tile 5: Data cleaning

- **Ppt**
- Data Quality 7

  -error

  -artefact

- Data Gathering 10
  - Preemptive
  - Retrospective
- Data Delivery 12
- Data Storage 14
- Data Integration 15
- Data Retrieval 17
- Data Mining and Analysis 18
- Conventional measures of data quality 23
- Data quality constraints 26

- Data quality constraints 27
- Methods for data cleaning 28
- Tools for Data Cleaning 32
- https://docs.google.com/presentation/d/1W7xX5-ys3MKYb4e6ycHLaURCd2_jw79jThLHGzjqSxk/edit#slide=id.g45192cef99_0_0

- **Docx**

  -What is “data quality”?
  - Consequences of poor quality data
  - Types of data errors
  - Data Errors in the Gathering Phase
  - Data Errors in the Storage Phase
  -Data Errors in the Processing Phase
  - Data Errors in the Analytics Phase
  - Measuring Data Quality
  https://docs.google.com/document/d/1JXHXwbhYyq2aOs8u7hhZfoyDdNc5mq_8_--aZYYTplg/edit#heading=h.xrr3hzmsua07

Docx-2

- Tools for Data Cleaning
- Tips on Data Cleaning
https://docs.google.com/document/d/16QhHykV31AwEMe7wL5qW6geW3Jy-1ddXUJXd-4uz6dY/edit#heading=h.j2exi2hd5f0n

Week 6

Tile 6: Communication

- **Ppt**
  - Graphic Communication 9
  - Graphic Communication: Stages of Understanding 11
  - Graphic Communication Goals 12
  - Goals: Persuasion 15
  - Goals: Education 18
  - Goals: Entertainment 19
  - Kirk’s principles of Good Data Visualisation 22
  https://docs.google.com/presentation/d/13RtXAsP9GArlHvZF4z_BTcvRIW_TFMz4CDEURSlghcE/edit

- **Docx**

  - What makes a “good” visualisation?
  https://docs.google.com/document/d/15gwfXToWNQnhGXZEsyKvjfFNbGOVDcFd5Vo_xGJ4A00/edit#heading=h.llin3unmr4ra

- **Docx -2**

  - Five Good Things to Know

1. Avoid pie charts
2. Don't use 3D effects in a 2D medium
3. Be careful about Axes
4. Watch out for distorted Area
5. Distraction and Clutter -- "Remove to Improve"

https://docs.google.com/document/d/1EbYuwsqqoANr1w4E8QeTCqeu4veo2EIXaeWfiIcZXz0/edit#heading=h.5ky0rde77jfr

Week 7

Tile 7: Encoding data

- **Ppt**

  - What is a graph? (or chart) 6
  - Data representation: Marks 9
  - Data representation: Attributes 13
  - Chart types 15
  - Chart diagram 27

  https://docs.google.com/presentation/d/1uIJ9tAVNalmXJV32KnxTL95aAoWeOEzQZrfeoelgSqY/edit

- **Docx-1**

  - Encoding data

  -Marks

  -Attributes

  - Perceptual ranking

  https://docs.google.com/document/d/11JRymp7kJe427xcz9NQ_ehXQf2hcwB2jScLWSTzoNCM/edit#heading=h.mgqdnuvk62fp

- **Docx-2**

  - A curated selection of charts for CA682

  - Categories of charts – CHRTS

https://docs.google.com/document/d/1gn1IQtqGknfJd5RR2qQPGh7ZeeJXlP8H85MdnnGnsy4/edit#heading=h.ldphmup34bbg

Week 8

Tile 8: Designing data-driven visualisations

- **PPT-** Data Visualiation III - human vision

  Graphic Communication: Stages of Understanding 16

  Binocular vision 19

  Depth Cues - Binocular Vision 21

  Depth Cues – Monocular 22

  Attention - Searchlight model 23

  Parallel Processes vs Serial Processes 26

  Pre-attentive processing 27

  https://docs.google.com/presentation/d/1X7sPY7sQPJ5sVCSMiab5lzAHOXPt2ofbVezfbyIP5p4/edit#slide=id.g25e7d590c63_0_167

- **Docx-** Preattentive Features

  - Form

  - Colour

  - Spatial position

  https://docs.google.com/document/d/1gcosfoduHt_VIGz3Q80taKxyMbBVOoDjTbWlfK3KcS8/edit#heading=h.a81epm1522q

- **PPT-** Data Visualiation III – design

  - Colour 4

  - Gestalt Laws 13

    Proximity

    Similarity

    Enclosure

    Closure

    Continuity

Connection

https://docs.google.com/presentation/d/1GyT8HYZ-t8Jv_nNce4x608GJ8rw_HXevU0sHMi4h-b0/edit#slide=id.ga243858335_0_152

- **Docx-** Visualisation Process

  - A Visualisation Process

  https://docs.google.com/document/d/1Z0eZOqCkNTq25AF2-M8lGSecDhs2pwkiIeTy6TsGf8c/edit#heading=h.nd4l6sixdq43

- **Docx-** Decluttering
  https://docs.google.com/document/d/1KUFjtk5p-eYgFaLqTauAorQASA6HQ6BwxxlwlgfFh48/edit#heading=h.k8merlahh190

- **Docx-** How to use colour

  - Some terms

    Hue

    Saturation

    Luminance

  - What about branding?

  - Choosing colours

  https://docs.google.com/document/d/119nETfGiWIy8VGcJhOgy-Zdb6K9JxzIx8RbmDfeuoAc/edit#heading=h.y12x76eq3auz

- **Docx-** Gestalt Theory

  Proximity

  Similarity

  Enclosure

  Closure

  Continuity

  Connection

  https://docs.google.com/document/d/1ZLC4Pj06YK4eWROzRNY4uNttY4CvupOil0b579gXwb8/edit

Data Visualization Checklist
https://loop.dcu.ie/pluginfile.php/5210580/mod_resource/content/1/DataVizChecklist_May2016.pdf

Week 9

Tile 9: Storing data

- **PPT-** Data Management
  - Database management tools 5
  - Data storage approaches 8

    Relational (traditional & modern)
    Column
    MPP, Data Warehouse
    NoSQL

    Big Data (MapReduce, Hadoop - HDFS)
  - OLAP, OLTP, DW ?? 12
  - So what is Map/Reduce 18
  - What about elasticsearch and ELK? 21
    https://docs.google.com/presentation/d/171GEKSVv_MVrXsZL-n5xRfFDQtN0EWAwDoluREwS43o/edit#slide=id.gaedf2bd6da_0_161

- **Docx-** Data Management Tools
  - data storage methods
  - data storage approaches
  - Relational Databases
  - Column Databases
  - Data Warehouses (DW) and Massively Parallel Processing (MPP)
  - NoSQL
  - Key-Value
  - Big Data Storage
  - Hadoop Distributed File System (HDFS)

  https://docs.google.com/document/d/1yLnf0AZ1XHlMyDAwwuC-lxog3TciiY_cT0c1c9H1w8Y/edit#heading=h.7ywankfitdnx

Week 10

Tile 10: Data protection and privacy

- **PPT- GDPR**

-GDPR? 3
- Data Protection v. Privacy Rights 6
- Personal Data 7
- Data Processing and Profiling 10
- Data Protection Principles I 15
- Legal challenges and risks 17
- Case-study I in research domain 18
https://drive.google.com/file/d/1pzMdIeNv8pVJZCel6a7aLoGaj6Wb1qVs/edit

- **Docx-** How your phone tracks your every move
  - Tracking device
  - Pattern recognition
  - The social network
  - What's in the data?

  https://www.abc.net.au/news/2015-08-17/metadata-retention-privacy-phone-will-ockenden/6694152?nw=0

Book reference links

https://www.getty.edu/publications/intrometadata/