CSC1143/CSC1175 Data Management & Visualisation

01 Introduction

Suzanne Little
suzanne.little@dcu.ie

Today

- Module outline
- Assessment
- Practicals
- Resources and support

- Who are you?
- What about data?

- A Data Analytics Pipeline
- Why do we visualise data?
- What's next?

Objectives for today:

- Get an overview of topics in DMV
- Get to know me
- Get to know a little about each other
- Ask me questions
- Experiment with technology :-)

Questions? I'll periodically stop but you can use the Q&A link at the top of the slides. Please put your name (first is fine) at the end of your question.

Housekeeping

Be polite and respectful to your fellow students and to me (have your phone on silent, stay quiet when I'm talking)

Keep the back 2 rows and the aisle seats free until after 4:15pm

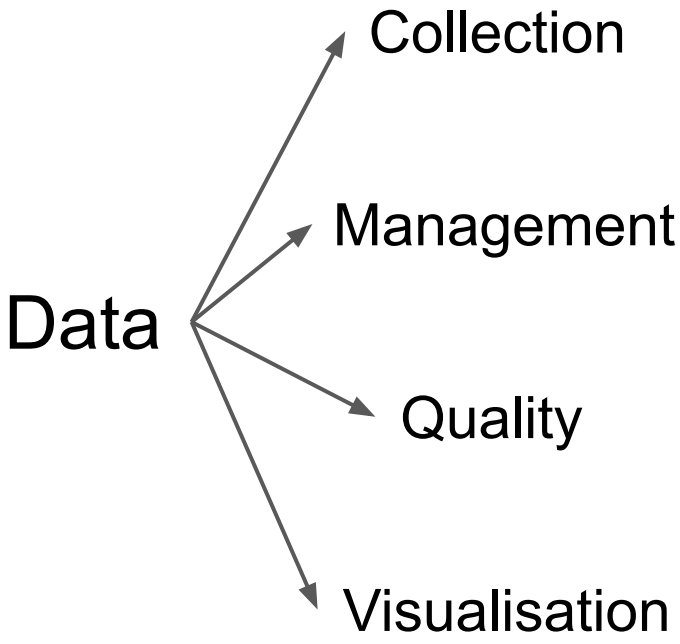
If an alarm sounds follow the exit signs outside to the assembly point

No food or drink allowed in the lecture theatre

I'll generally give you a 10 min break part way through

About me

Module Content Outline



Outcomes

1. Analyse the requirements of applications handling large datasets.
2. Demonstrate an ability to efficiently structure a large dataset.
3. Implement data quality measures.
4. Identify and implement appropriate data visualization techniques.

DMV module specifics

Fully in-person delivery → not recorded or broadcast

7.5 credits

→ Thu 4-5:50pm, lecture, discussions, case studies, activities

→ Thu 6-6:50pm, labs, practical work, programming, working on assignment

→ Tutors will be available in the labs to lend a hand

→ plus self study and assignment work

Assessment: 25% assignment, 75% exam

Resit Category 3: if you fail then you can only retake the exam (75%), not the assignment

Resources and Support

Loop (Moodle)

- main source of communication, notes, resources
- check your student emails! (@mail.dcu.ie)
- links to shared materials on Google Drive (must be logged in @mail.dcu.ie)
- CA682? CA682A? D? E? → CSC1143 (MCM) & CSC1175 (Law & Gov)

DCU Library

- Online ebooks (see Loop for a list and links to resources)

Myself

- Email is best for contacting me (suzanne.little@dcu.ie)
- Please use your @mail.dcu.ie address if possible & start the subject with **[DMV] or [CSC1143] or [CSC1175]**

Resources

There is no set textbook. Reference material will be available online.

Two books (available via the DCU library) that address Data Visualisation well:

- “Data Visualisation”, Andy Kirk (2016), Sage Publishers → 2012 version as ebook
- “Storytelling with Data”, Cole Nussbaumer Knaflic (2015), Wiley

Also: “The Data Science Handbook”, Field Cady (2017), ebook

Available on loop: slides, some notes, links to readings, practical exercises, revision quizzes, discussion topics

Access problems during week 1

I’m making the first weeks materials available at https://github.com/suzannelittle/dcu-dmv

If you aren’t on the class mailing list you won’t have permission to view Google Files (slides, documents). Please be patient and don’t request access (spams my inbox!)

Use the github.com repository for now

Assessment -- more detail later

Assignment 25%

- Create a complex data visualisation and present using a short screencast (<5mins)
- Due Friday November 29th via upload to loop
- More information later - will be in pairs

Exam 75%

- Multiple choice and short answer questions
- Past papers and examples are/will be available online
- On lab machines, during the exam session

Questions?

Today

- Module outline
- Assessment
- Practicals
- Resources and support

- Who are you?

- What about data?
- A Data Analytics Pipeline
- Why do we visualise data?
- What's next?

Getting to know you

https://vevox.app

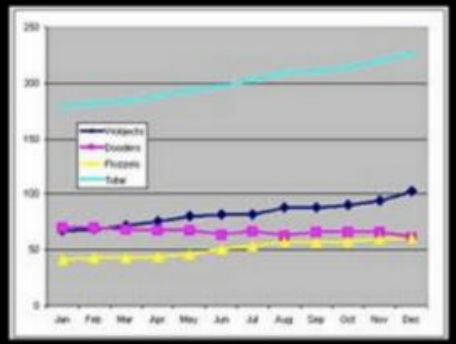
194-124-435

Data Scientist

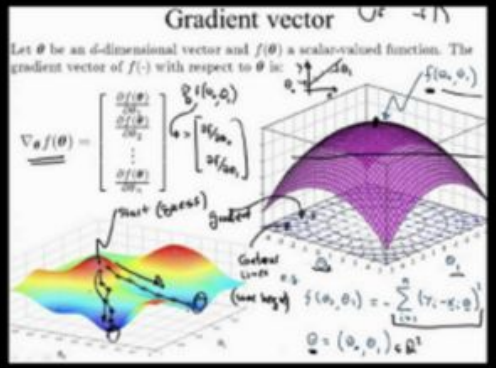
What my friends think I do

What my mom thinks I do

What society thinks I do



What my boss thinks I do



What I think I do

What I actually do

Housekeeping

Be polite and respectful to your fellow students and to me (have your phone on silent, stay quiet when I'm talking)

Keep the back 2 rows and the aisle seats free until after 9:15am

If an alarm sounds follow the exit signs outside to the assembly point

No food or drink allowed in the lecture theatre

I'll generally give you a 10 min break part way through

Today

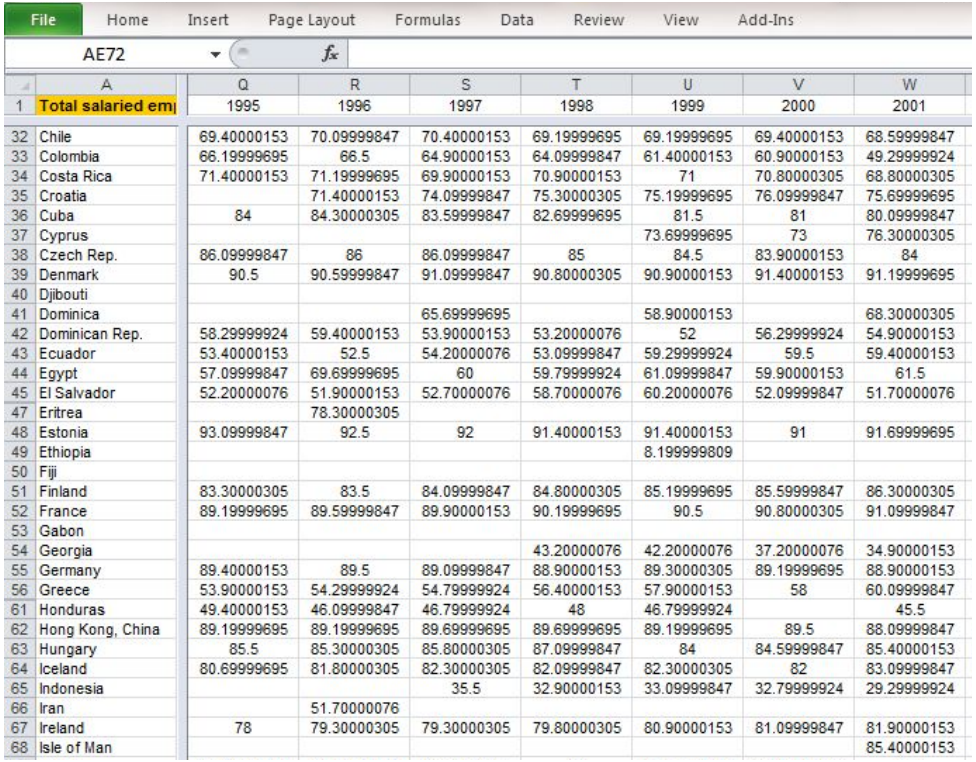
- What about data?
- A Data Analytics Pipeline
- Why do we visualise data?
- What's next?

Question: Is it “data is” or “data are”?

Data is or Data are ?
https://www.theguardian.com/news/datablog/2010/jul/16/data-plural-singular

Discussion: What is data?
How many potential data sources can you think of?
https://vevox.app
121-595-356

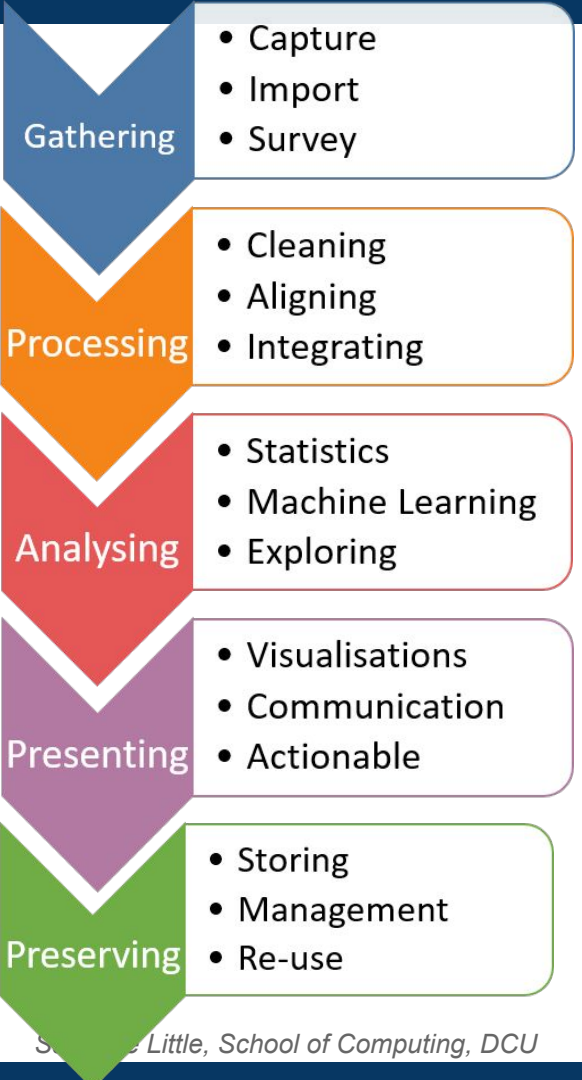
Data



	A	Q	R	S	T	U	V	W
1	Total salaried em	1995	1996	1997	1998	1999	2000	2001
32	Chile	69.40000153	70.09999847	70.40000153	69.19999695	69.19999695	69.40000153	68.59999847
33	Colombia	66.19999695	66.5	64.90000153	64.09999847	61.40000153	60.90000153	49.29999924
34	Costa Rica	71.40000153	71.19999695	69.90000153	70.90000153	71	70.80000305	68.80000305
35	Croatia		71.40000153	74.09999847	75.30000305	75.19999695	76.09999847	75.69999695
36	Cuba	84	84.30000305	83.59999847	82.69999695	81.5	81	80.09999847
37	Cyprus					73.69999695	73	76.30000305
38	Czech Rep.	86.09999847	86	86.09999847	85	84.5	83.90000153	84
39	Denmark	90.5	90.59999847	91.09999847	90.80000305	90.90000153	91.40000153	91.19999695
40	Djibouti							
41	Dominica			65.69999695		58.90000153		68.30000305
42	Dominican Rep.	58.29999924	59.40000153	53.90000153	53.20000076	52	56.29999924	54.90000153
43	Ecuador	53.40000153	52.5	54.20000076	53.09999847	59.29999924	59.5	59.40000153
44	Egypt	57.09999847	69.69999695	60	59.79999924	61.09999847	59.90000153	61.5
45	El Salvador	52.20000076	51.90000153	52.70000076	58.70000076	60.20000076	52.09999847	51.70000076
47	Eritrea		78.30000305					
48	Estonia	93.09999847	92.5	92	91.40000153	91.40000153	91	91.69999695
49	Ethiopia					8.199999809		
50	Fiji							
51	Finland	83.30000305	83.5	84.09999847	84.80000305	85.19999695	85.59999847	86.30000305
52	France	89.19999695	89.59999847	89.90000153	90.19999695	90.5	90.80000305	91.09999847
53	Gabon							
54	Georgia				43.20000076	42.20000076	37.20000076	34.90000153
55	Germany	89.40000153	89.5	89.09999847	88.90000153	89.30000305	89.19999695	88.90000153
56	Greece	53.90000153	54.29999924	54.79999924	56.40000153	57.90000153	58	60.09999847
61	Honduras	49.40000153	46.09999847	46.79999924	48	46.79999924		45.5
62	Hong Kong, China	89.19999695	89.19999695	89.69999695	89.69999695	89.19999695	89.5	88.09999847
63	Hungary	85.5	85.30000305	85.80000305	87.09999847	84	84.59999847	85.40000153
64	Iceland	80.69999695	81.80000305	82.30000305	82.09999847	82.30000305	82	83.09999847
65	Indonesia			35.5	32.90000153	33.09999847	32.79999924	29.29999924
66	Iran		51.70000076					
67	Ireland	78	79.30000305	79.30000305	79.80000305	80.90000153	81.09999847	81.90000153
68	Isle of Man							85.40000153

log files
social media content
photographs
microblogs
surveys
news
cctv video
movies
television
sales records
clicks
adwords
statistics
audio recordings
playlists
search terms
sensors
pedometer/activity monitor
spectrographs
microscopy
genomes
numbers

Data is collected information (a working definition)



Many sources: websites, user surveys, sensors, legacy databases, and more!

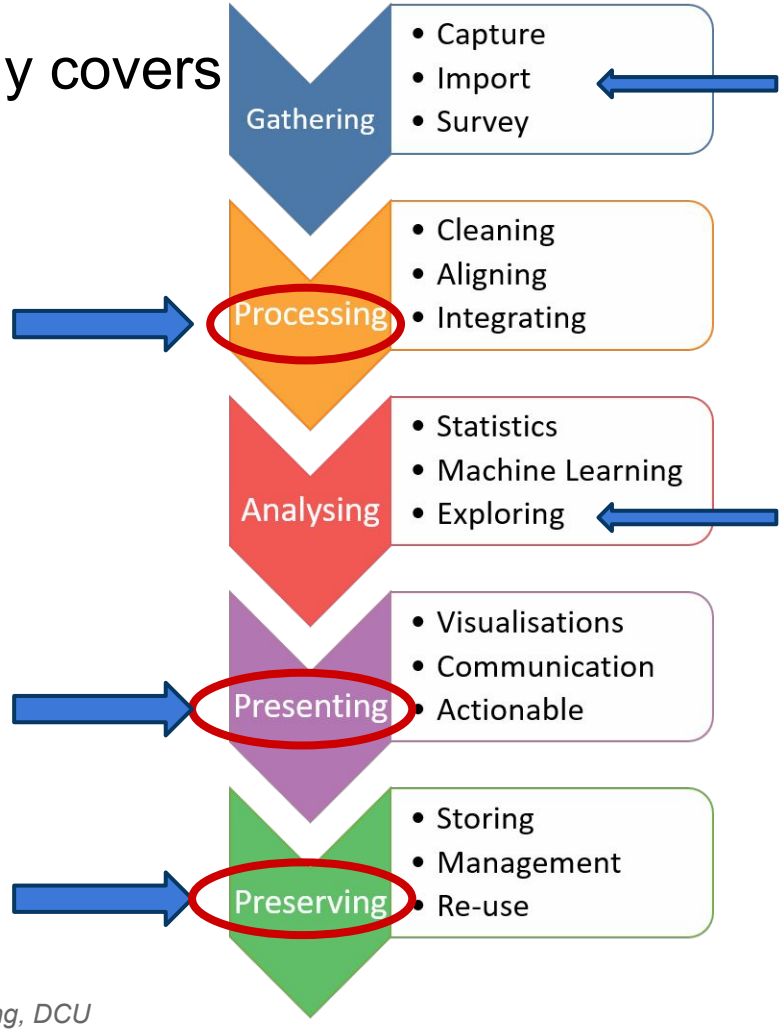
computers are simple,
people are complicated

Number crunching, discovering patterns

What is your message?
Who is your audience?

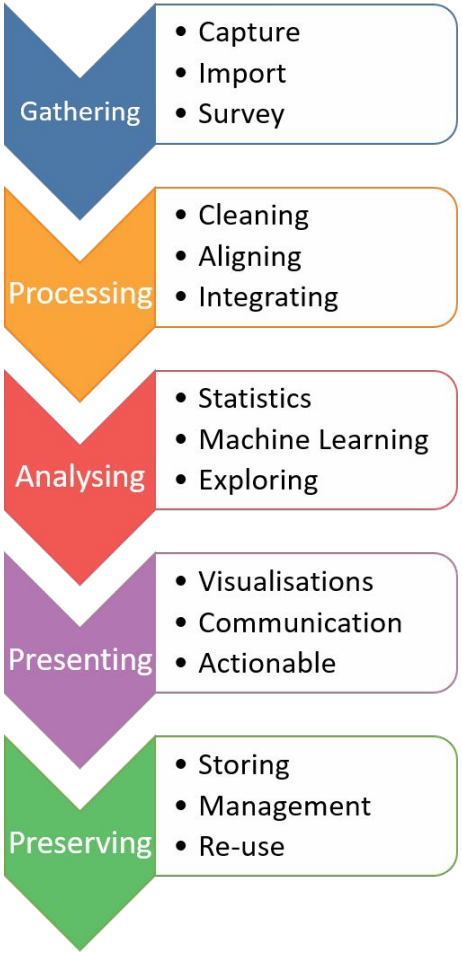
Data storage & indexing, recording processes (provenance)

CSC1143 mostly covers



Topics plan (subject to change!)

1. What is data?
2. Describing data
3. Finding data
4. Cleaning data
5. Communication
6. Encoding data
7. Designing data-driven visualisations
8. Managing and Storing data + Big Data
9. Data protection & privacy
10. Final wrap up and exam information



What background do I need to have?

Or what technologies will you be teaching?

Data Management & Visualisation module is *technology agnostic*

That is, you learn the fundamental principles and apply them using a range of tools.

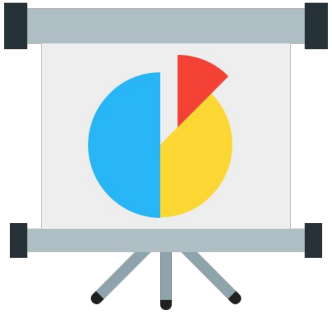
Labs will include exercises using Python, Tableau, OpenRefine, Spreadsheets

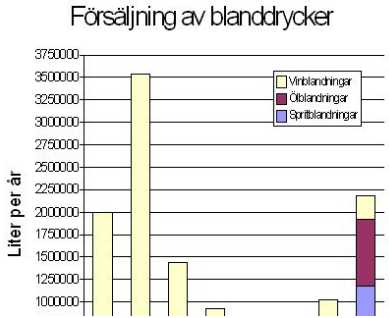
Your assignment can be completed using many different tools. The exam won't specify which tool to use.

Complete the class background skills survey and we'll discuss more next week

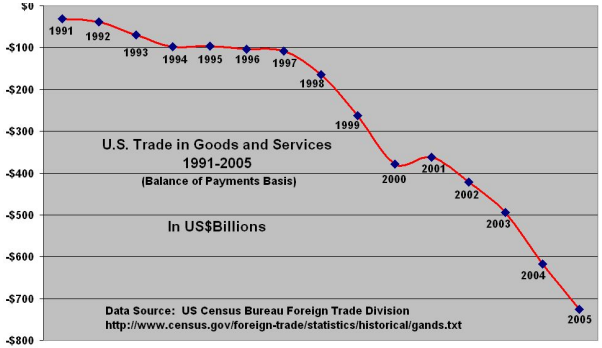
Questions?

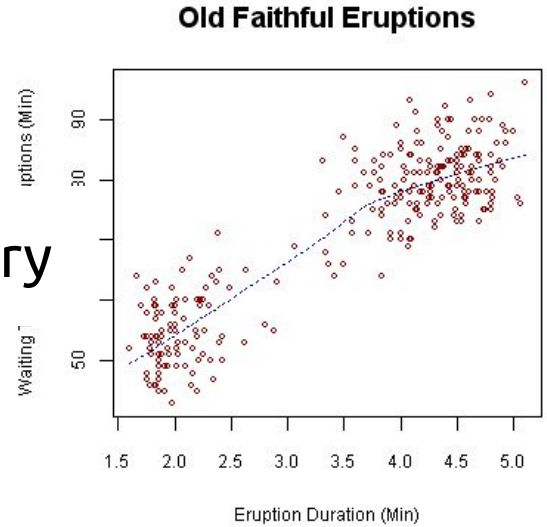
Why do we visualise data?

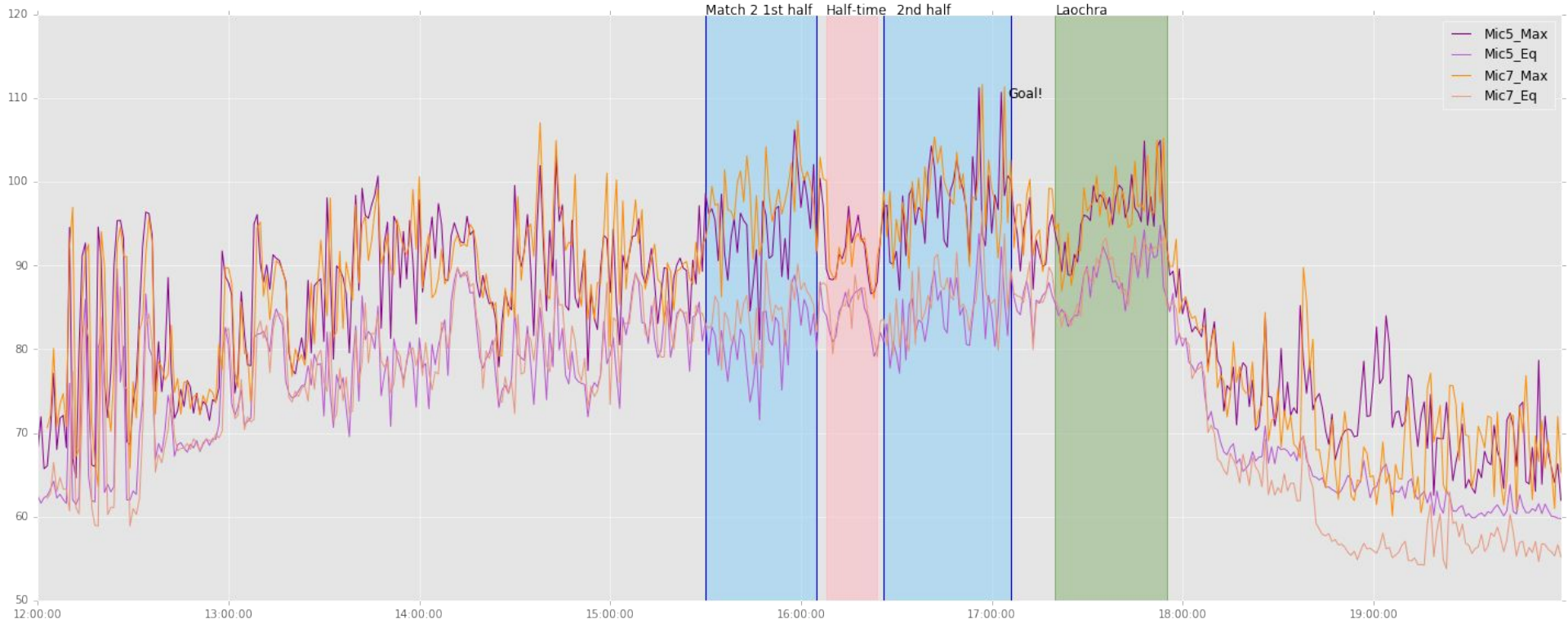




Exploratory vs Explanatory

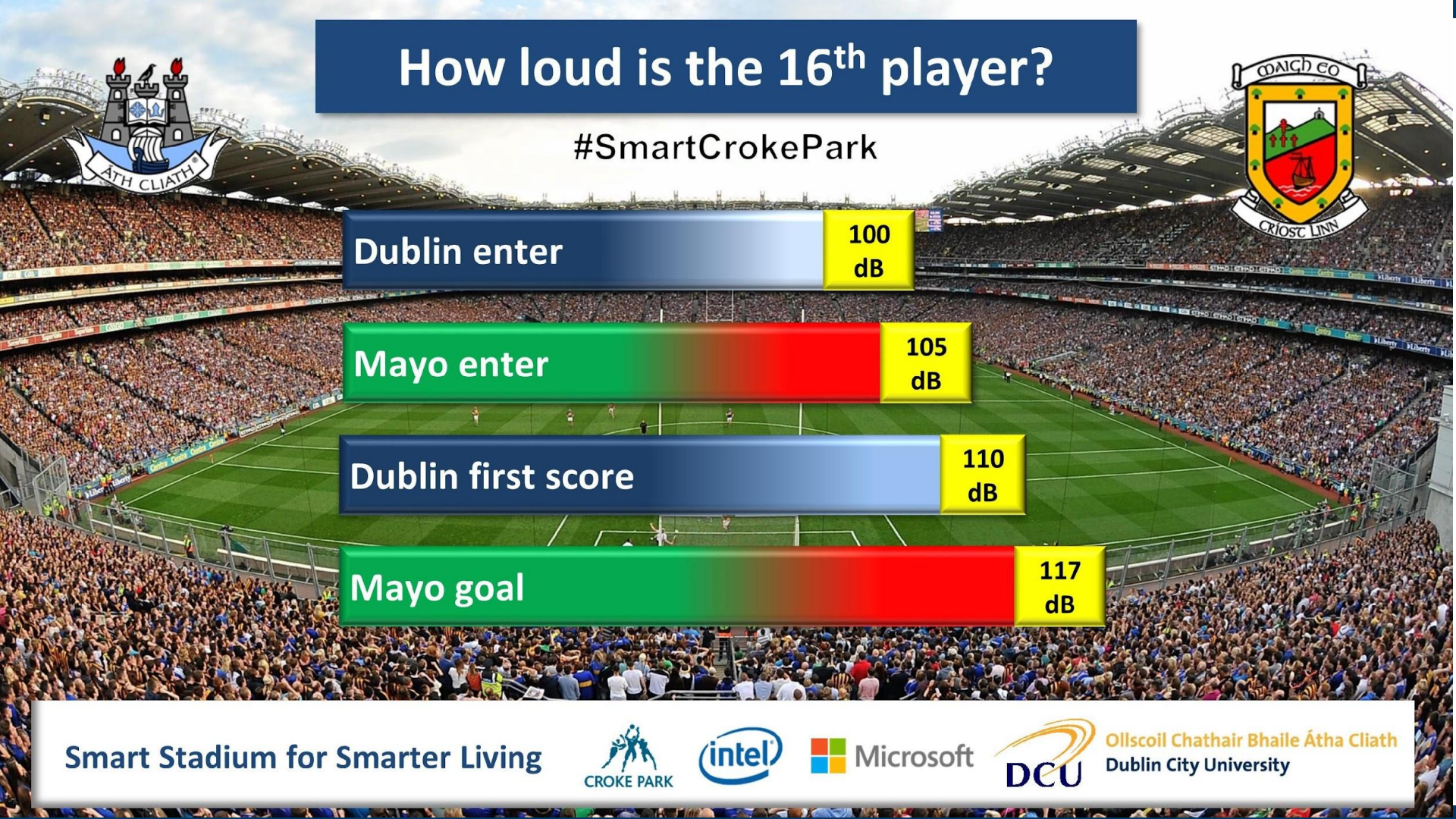





Match 2 1st half
Half-time
2nd half
Laochra
Goal!
Mic5_Max
Mic5_Eq
Mic7_Max
Mic7_Eq
120
110
100
90
80
70
60
50
12:00:00
13:00:00
14:00:00
15:00:00
16:00:00
17:00:00
18:00:00
19:00:00

How loud is the 16th player?

#SmartCrokePark



Smart Stadium for Smarter Living

CROKE PARK

intel

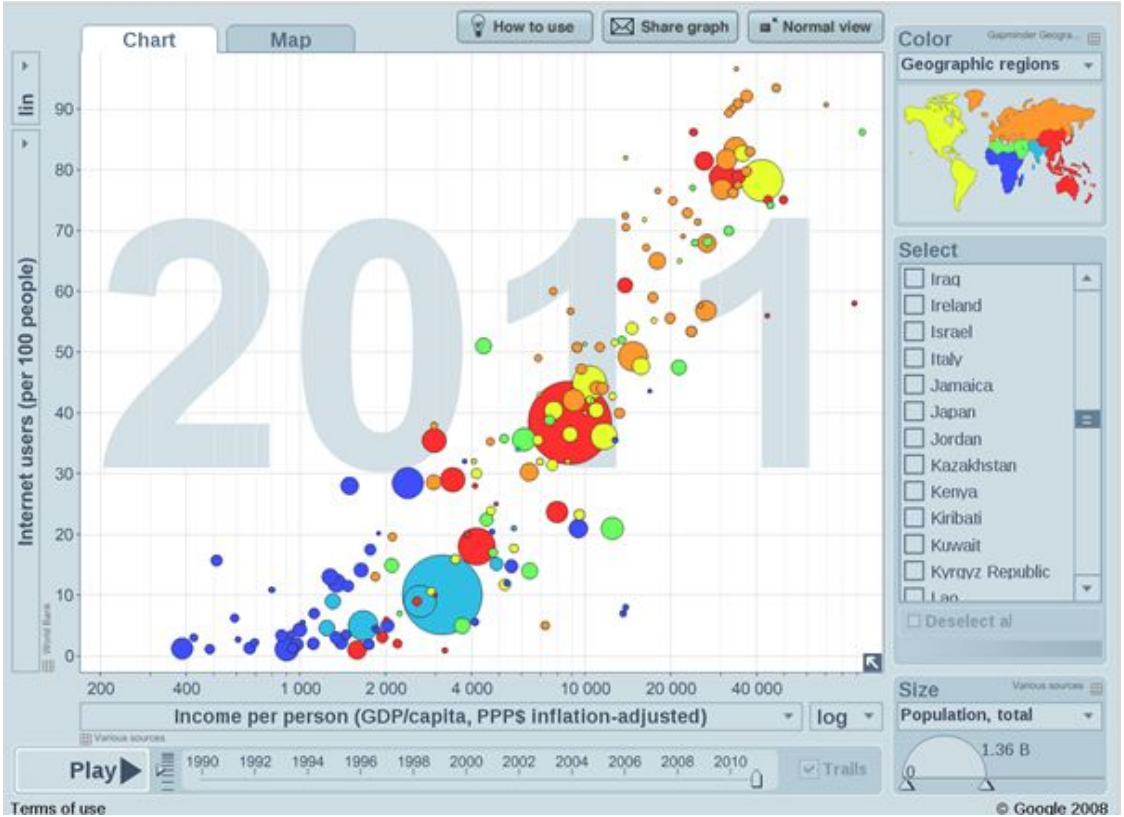
Microsoft

DCU Ollscoil Chathair Bhaile Átha Cliath Dublin City University

Questions?

Hans Rosling
See TED talks on economics
https://youtu.be/hVimVzgtD6w

Data Visualisation: Access to the Internet



www.gapminder.org

Today's Lab (6-7pm)

LG25: Family name A-G
LG26: Family name H-M
L128: Family name N-R
L125: Family name S-Z

1. Complete the background skill survey - https://forms.gle/BTtbEEeSN9VCdPWw7
2. Explore gapminder and answer the questions
3. Revision Quiz on loop (no marks, just feedback)

Next week

Resources will be shared with you - linked from Loop

For this week I'll also put links to the slides and lab information on https://github.com/suzannelittle/dcu-dmv/

Before next week:

- complete the background skills survey
- upload your discoveries from gapminder
- review documents on loop (or check gitlab for links)

Questions?

Links & Resources

GapMinder - www.gapminder.org

Hans Rosling TED - https://youtu.be/hVimVzgtD6w

Hans Rosling BBC - https://youtu.be/jbkSRLYSojo

→ CSC1175 students could you wait please ...