A Tutorial on Data Storytelling for Learning Analytics Dashboards
Organizers

Vanessa Echeverria
ESPOL University

Lu Lawrence
CMU

Yi-Shan Tsai
Monash University

Roberto Martinez-Maldonado
Monash University

Shaveen Singh
Monash University

Gloria Fernandez-Nieto
UTS
## Schedule and Activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Start time; Time for activity</th>
</tr>
</thead>
</table>
| Committee introduction  
  Participants introduce themselves in the chat  
  (Name, Pronouns, Title, Location)                                  | 0:00; 5 minutes               |
| Lecture explaining data storytelling principles                          | 0:05; 30 minutes              |
| Introduce first half of activity                                         | 0:35; 10 minutes              |
| Social Break Gathertown                                                 | 0:45; 15 minutes              |
| Come back and split into breakout rooms by groups                       | 1:00; 5 minutes               |
| First half of activity – Context                                        | 1:05; 20 minutes              |
| **Reflection as a group**  
  *Each group will present for 1-2 minutes about the data, their audience, and the intention*  
                                              | 1:25; 15 minutes              |
| Introduce second half of activity                                       | 1:40; 20 minutes              |
| Second half of activity – Visuals, declutter, attention, and design     | 2:00; 30 minutes              |
| **Reflection and discussion**  
  *Each group presents their experience*  
  *Roberto leads discussion about DS principles, automating this process, and role of the designer*  
                                              | 2:30; 30 minutes              |
Educational data storytelling

Dr. Vanessa Echeverria
ESPOL University
Exploratory visualizations
Please approve the hire of 2 FTEs
to backfill those who quit in the past year

Ticket volume over time

2 employees quit in May. We nearly kept up with incoming volume in the following two months, but fell behind with the increase in Aug and haven't been able to catch up since.
Please approve the hire of 2 FTEs
to backfill those who quit in the past year

Ticket volume over time

2 employees quit in May. We nearly kept up with incoming volume in the following two months, but fell behind with the increase in Aug and haven’t been able to catch up since.

Explanatory visualizations
Exploratory vs **explanatory** visualizations

Hunting for pearls in oysters

Communicating a specific story

What is data storytelling?
Storytelling

History → Words → Trust

Experience → Memory
Find stories in data

Not just about data analysis
Data storytelling elements
Data comics to tell stories

The Emerging Genre of Data Comics

Benjamin Bach, Harvard University
Nathalie Henry Riche, Microsoft Research
Sheelagh Carpendale, University of Calgary
Hanspeter Pfister, Harvard University

- Engage
- Break complexity
- Guide

https://datacomics925658343.wordpress.com/portfolio/co2-footprint/
Data comics to tell stories

We quantify the structural properties of these graphs by their **characteristic path length** \( L(p) \) and **clustering coefficient** \( C(p) \).

\( L(p) \) measures the typical separation between two vertices (a global property). \( C(p) \) measures the cliquishness of a typical neighbourhood (a local property).

\( L \) is defined as the number of edges in the shortest path between two vertices

\( C \) is defined as follows. Suppose that a vertex \( v \) has \( k_v \) neighbours.

Then at most \( k_v (k_v - 1) / 2 \) edges can exist between them. (This occurs when every neighbor of \( v \) is connected to every other neighbour of \( v \).)

Let \( C_v \) denote the fraction of these allowable edges that actually exist. Define \( C \) as the average of \( C_v \) over all vertices.

For friendship networks, these statistics have intuitive meanings: \( L \) is the average number of friendships in the shortest chain connecting two people. \( C_v \) reflects the extent to which friends of \( v \) are also friends of each other; and thus \( C \) measures the cliquishness of a typical friendship circle.
A Start-Up Story
Sam, Charles and Francis are three friends...

... founding their own start-up...
Creative & Co.

... and looking for collaborators.

They found a company working on a similar idea...

... with which they kept increasing collaboration.
Educational data storytelling

Communicate insights to support teaching and learning practices
Exploratory visualisation about group’s performance
Explanatory visualisation about group’s performance

High-performing Team
The team was able to determine all main entities and all main relationships

Well done! The team recognised important entities and relationships.
High-performing Team

The team was able to determine all main entities and all main relationships.

The team’s solution improved a lot in a short period of time.

90% “Scholarship - Student” relationship

70% “Scholarship - Deposit” relationship

45% “Scholarship” entity

35% “Student” entity

21% “Student” and “Scholarship” were created as attributes

Well done! The team recognised important entities and relationships.
Explanatory visualisation about group’s performance

High-performing Team
The team was able to determine all main entities and all main relationships

Decluttering

The team’s solution improved a lot in a short period of time

“Scholarship - Student” relationship

90%

Well done!
The team recognised important entities and relationships.
Explanatory visualisation about group’s performance

High-performing Team
The team was able to determine all main entities and all main relationships

Decluttering
“Student” and “Scholarship” were created as attributes

Well done!
The team recognised important entities and relationships.
Explanatory visualisation about group’s performance

High-performing Team
The team was able to determine all main entities and all main relationships

Selected data points
- 21% “Student” entity
- 35% “Student” and “Scholarship” were created as attributes
- 45% “Scholarship” entity
- 70% “Scholarship - Deposit” relationship
- 90% “Scholarship - Student” relationship

Decluttering

Well done! The team recognised important entities and relationships.
Explanatory visualisation about group’s performance

High-performing Team
The team was able to determine all main entities and all main relationships

Text explaining trends
The team’s solution improved a lot in a short period of time

Selected data points
35% “Student” entity
21% “Student” and “Scholarship” were created as attributes

Decluttering

Well done!
The team recognised important entities and relationships.
High-performing Team
The team was able to determine all main entities and all main relationships

Text explaining trends
The team’s solution improved a lot in a short period of time

Selected data points
35%
“Student” entity
45%
“Scholarship” entity
70%
“Scholarship - Student” relationship
90%
“Scholarship - Deposit” relationship

Decluttering
21%
“Student” and “Scholarship” were created as attributes

Assessment narratives
90%
Well done!
The team recognised important entities and relationships.
Connecting the dots...
Equal participation

Highlight where there are not equal participation

Text label, colours, data points
Educational data storytelling
Communicate insights to support teaching and learning practices
Team 2

After the patient lost consciousness, the team reacted slowly.

It is recommended to call the doctor earlier, preferably during the observations.
Turning data into stories

Data → Analysis → Insight → Story
Turning data into stories

Data Analysis Insight Story
Turning data into stories

With the help of data storytelling techniques
Turning findings into stories

Freytag’s Pyramid & Data Stories

- **Introduction** = Set-up/Background
- **Inciting Incident** = Deviation from expected
- **Rising Action** = Supporting Facts
- **Climax** = Main Insight
- **Resolution** = Recommendation
- **Conclusion** = Next Steps

Character Growth = Increase in audience awareness
How to **tell stories** with data

1. Understand the context
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it
5. Think like a designer
6. Tell a story
1. Understand the context

1. **WHO**
   - is your **AUDIENCE**?
   - BE SPECIFIC!

2. **WHAT**
   - DO YOU NEED them to do?
   - BE EXPLICIT!

3. **HOW**
   - WILL DATA HELP make your point?
   - BE DISCERNING!

[Image showing diagram of questions: What is their relationship to you? What motivates them? What keeps them up at night? Don't assume they will connect the dots!]

20% of children had a traditional stay-at-home mom in 2012, compared to 41% in 1970.

If you want to communicate one or two values.
### Tables

Avoid heavy borders to improve the legibility of your table.
Scatter plot

If you want to show the relationship between two things.
Line graphs

If you want to plot continuous data and look for trends.
Bar graphs

Easy to read and quickly see which category is the biggest.
Horizontal bar chart

Categorical data and names of categories are long.
Let’s review some examples....
Supplier Market Share

- Supplier A
- Supplier B
- Supplier C
- Supplier D

The donut chart

arc A
arc B
Avoid these types of visuals!
how many issues were there in January and February?
how many issues were there in January and February?

Avoid these types of visuals!
Source: http://www.thefunctionalart.com/2015/02/redesigning-circular-timeline.html
Avoid this!

Source: http://www.thefunctionalart.com/2015/02/redesigning-circular-timeline.html
What is confusing in this graph?
What is confusing in this graph?

Left y-axes meaning

Right y-axes meaning

Is it a time series?
nice use of the slopegraph
3. Eliminate clutter

Visual elements that take up space and don’t aid our understanding.
Decluttering step by step

1. Remove chart borders
2. Remove gridlines
3. Remove data markers
4. Clean up axis labels
5. Label data directly
6. Leverage consistent color
Decluttering step by step

Before & after

Shoppers Begins Shopping for Holidays

More women start their holiday shopping early

% OF TOTAL

Before Sep  Sep  Oct  Nov  Dec: first 2 weeks  Dec: last 2 weeks

9%  16%  20%  43%  18%  4%
5%  8%  21%  40%  13%  2%
4. Focus attention where you want it
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No preattentive attributes</strong>&lt;br&gt;No preattention attributes are used in the text.</td>
<td>No preattention attributes are used in the text.</td>
</tr>
<tr>
<td><strong>Bold</strong></td>
<td>The word &quot;Bold&quot; is in bold.</td>
</tr>
<tr>
<td><strong>Color</strong></td>
<td>The words &quot;Great Products&quot; are in a different color.</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>The word &quot;are&quot; is larger than the surrounding text.</td>
</tr>
<tr>
<td><strong>Separate spatially</strong></td>
<td>The words &quot;in the billing office&quot; are in a separate spatial location.</td>
</tr>
<tr>
<td><strong>Outline (enclosure)</strong></td>
<td>The word &quot;sent&quot; is enclosed with a box.</td>
</tr>
<tr>
<td><strong>Underline (added marks)</strong></td>
<td>The word &quot;quick&quot; is underlined with an added mark.</td>
</tr>
</tbody>
</table>
PREATTENTIVE ATTRIBUTES

SIGNAL WHERE to LOOK and create VISUAL HIERARCHY to help ease the processing of information.

<table>
<thead>
<tr>
<th>ORIENTATION</th>
<th>SHAPE</th>
<th>LINE LENGTH</th>
<th>LINE WIDTH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SIZE</th>
<th>CURVATURE</th>
<th>ADDED MARKS</th>
<th>ENCLOSURE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HUE</th>
<th>INTENSITY</th>
<th>SPATIAL POSITION</th>
<th>MOTION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What about the color?
Some guidelines in data visualization:

1. Colour should add information:

![Bar chart showing Google tutorial searches for Python, JavaScript, and R.](chart1)

![Bar chart showing another set of Google tutorial searches for Python, JavaScript, and R.](chart2)
So, how should you use colour then?

1. Colour to differentiate
2. Colour to Explore
2. Types of colour schemes:

- Qualitative
- Sequential
- Diverging
3. Consider colour blindness:

https://www.color-blindness.com/coblis-color-blindness-simulator/
6. Tell a story
<table>
<thead>
<tr>
<th>Visual elements</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heading</td>
<td>Shows a straightforward message from the graph</td>
</tr>
<tr>
<td>Lines to emphasize</td>
<td>Emphasize relevant information</td>
</tr>
<tr>
<td>Key data points</td>
<td>Focus attention on key data</td>
</tr>
<tr>
<td>Decluttering</td>
<td>Remove elements that do not add value to the graph</td>
</tr>
<tr>
<td>Text labels</td>
<td>Add narratives to the graph</td>
</tr>
<tr>
<td>Shaded areas</td>
<td>Add context to a group of data points</td>
</tr>
</tbody>
</table>

*Echeverria et al. (2018)*
6. Tell a story

Students’ assessment results and total number of interactions with VLE resources module AAA

1. Interactions with VLE


2013J students got the highest average in their assessments and their VLE interactions were higher compared to other presentations. 2014J students got lower average in their assessments and their VLE’ interactions were smaller that 2013J students.

2. Assessment average results

- Presentation 2013J: 66.8 (1753), 66.2 (1756), 69.1 (1752), 69.3 (1754), 69.6 (1755)
- Presentaton 2014J: 69.2 (1753), 69.2 (1756), 70.3 (1752), 70.4 (1754), 70.6 (1755)
Break!

GatherTown link here!
Part 2: Hands on activity
Structure of the activity:

1. Understand the context
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it
5. Design!
1. Understand the context

Using:

https://analyse.kmi.open.ac.uk/open_dataset#about

Open University Learning Analytics dataset

This page introduces the anonymised Open University Learning Analytics Dataset (OULAD). It contains data about courses, students and their interactions with Virtual Learning Environment (VLE) for seven selected courses (called modules). Presentations of courses start in February and October - they are marked by “B” and “J” respectively. The dataset consists of tables connected using unique identifiers. All tables are stored in the csv format.
1. Understand the context - Who

Using:

https://analyse.kmi.open.ac.uk/open_dataset#about

Open University Learning Analytics dataset

This page introduces the anonymised Open University Learning Analytics Dataset (OULAD). It contains data about courses and students and their interactions with Virtual Learning Environment (VLE) for seven selected courses (called modules). Presentations of courses start in February and October - they are marked by “B” and “J” respectively. The dataset consists of tables connected using unique identifiers. All tables are stored in the csv format.
1. Understand the context - What

Using:

https://analyse.kmi.open.ac.uk/open_dataset#about

Open University Learning Analytics dataset

This page introduces the anonymised Open University Learning Analytics Dataset (OULAD). It contains data about courses, students and their interactions with Virtual Learning Environment (VLE) for seven selected courses (called modules). Presentations of courses start in February and October - they are marked by “B” and “J” respectively. The dataset consists of tables connected using unique identifiers. All tables are stored in the csv format.
1. Understand the context - What

1. Study students interaction with the VLE and
2. Identify relationship between the students VLE interactions and their outcomes (assessments).
1. Understand the context - How

1. What data do we have available?
   • e.g. interactions with a Virtual Learning Environment
1. Understand the context

<table>
<thead>
<tr>
<th>code_module</th>
<th>code_presentation</th>
<th>activity_type</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>dataplus</td>
<td>1843</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>forumng</td>
<td>175513</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>glossary</td>
<td>327</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>homepage</td>
<td>140345</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>oucollaborate</td>
<td>241</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>oucontent</td>
<td>267482</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>resource</td>
<td>7365</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>subpage</td>
<td>39624</td>
</tr>
<tr>
<td>AAA</td>
<td>2013J</td>
<td>url</td>
<td>15754</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>dataplus</td>
<td>1555</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>forumng</td>
<td>161590</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>glossary</td>
<td>329</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>homepage</td>
<td>126375</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>oucollaborate</td>
<td>257</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>oucontent</td>
<td>255912</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>resource</td>
<td>7559</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>subpage</td>
<td>31958</td>
</tr>
<tr>
<td>AAA</td>
<td>2014J</td>
<td>url</td>
<td>12623</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>code_module</th>
<th>code_presentation</th>
<th>activity_type</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>forumng</td>
<td>811107</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>glossary</td>
<td>1906</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>homepage</td>
<td>291961</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>oucontent</td>
<td>1375</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>ouelluminate</td>
<td>1433</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>quiz</td>
<td>93805</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>resource</td>
<td>46552</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>sharedsubpage</td>
<td>59</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>subpage</td>
<td>79867</td>
</tr>
<tr>
<td>BBB</td>
<td>2013B</td>
<td>url</td>
<td>19846</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>forumng</td>
<td>481790</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>glossary</td>
<td>1697</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>homepage</td>
<td>229807</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>oucollaborate</td>
<td>932</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>oucontent</td>
<td>13528</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>quiz</td>
<td>80105</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>resource</td>
<td>31932</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>sharedsubpage</td>
<td>57</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>subpage</td>
<td>53398</td>
</tr>
<tr>
<td>BBB</td>
<td>2013J</td>
<td>url</td>
<td>28530</td>
</tr>
</tbody>
</table>

**QUESTION 1:** How many different ways can you come up with to show this data?

**QUESTION 3:** Which of the visuals you imagined do you like best?
2. Choose an appropriate visual design
3. Eliminate clutter – in our context

Avoid repeating information
4. Focus attention where you need it

Use color, design elements, etc
4. Design!

Add image about using sticky notes to brainstorm designs and retain flexibility
Hands on activity

Structure for the activity is in Miro

One member of the organizing committee will check in with your group to help you get started and answer any questions
Practice Example - turn **this data** into a **data story**

Survey results: summer learning program on science

**PRE: How do you feel about doing science?**

<table>
<thead>
<tr>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excited</td>
<td>19</td>
</tr>
<tr>
<td>Interested</td>
<td>25</td>
</tr>
<tr>
<td>Ok</td>
<td>40</td>
</tr>
<tr>
<td>Not great</td>
<td>5</td>
</tr>
<tr>
<td>Boring</td>
<td>11</td>
</tr>
</tbody>
</table>

**POST: How do you feel about doing science?**

<table>
<thead>
<tr>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excited</td>
<td>38</td>
</tr>
<tr>
<td>Interested</td>
<td>30</td>
</tr>
<tr>
<td>Ok</td>
<td>14</td>
</tr>
<tr>
<td>Not great</td>
<td>6</td>
</tr>
<tr>
<td>Boring</td>
<td>12</td>
</tr>
</tbody>
</table>