

Chart Types

EDC101: Week 4

Dr. Calum Webb

Sheffield Methods Institute, the University of Sheffield
c.j.webb@sheffield.ac.uk

Sign in



Learning outcomes

What will I learn?

By the end of this week you will know:

- Key terminology for working with quantitative data
- Two different ways to help us think about which kind of data visualisation we should use:
 - The "classic" approach, where the type of data informs the visualisation
 - Some more modern approaches, where the data story helps inform the choice of the visualisation

How do I choose what kind of data visualisation is appropriate for my data?



Part I

Some key terminology



This is a dataset

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

Showing 1 to 15 of 144 entries, 22 total columns

Each **row** is an **observation**, e.g. a song, person, measurement at a specific time

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

Showing 1 to 15 of 144 entries, 22 total columns

Each **column** is a **variable**, e.g. the name of the song, the tempo, etc.

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

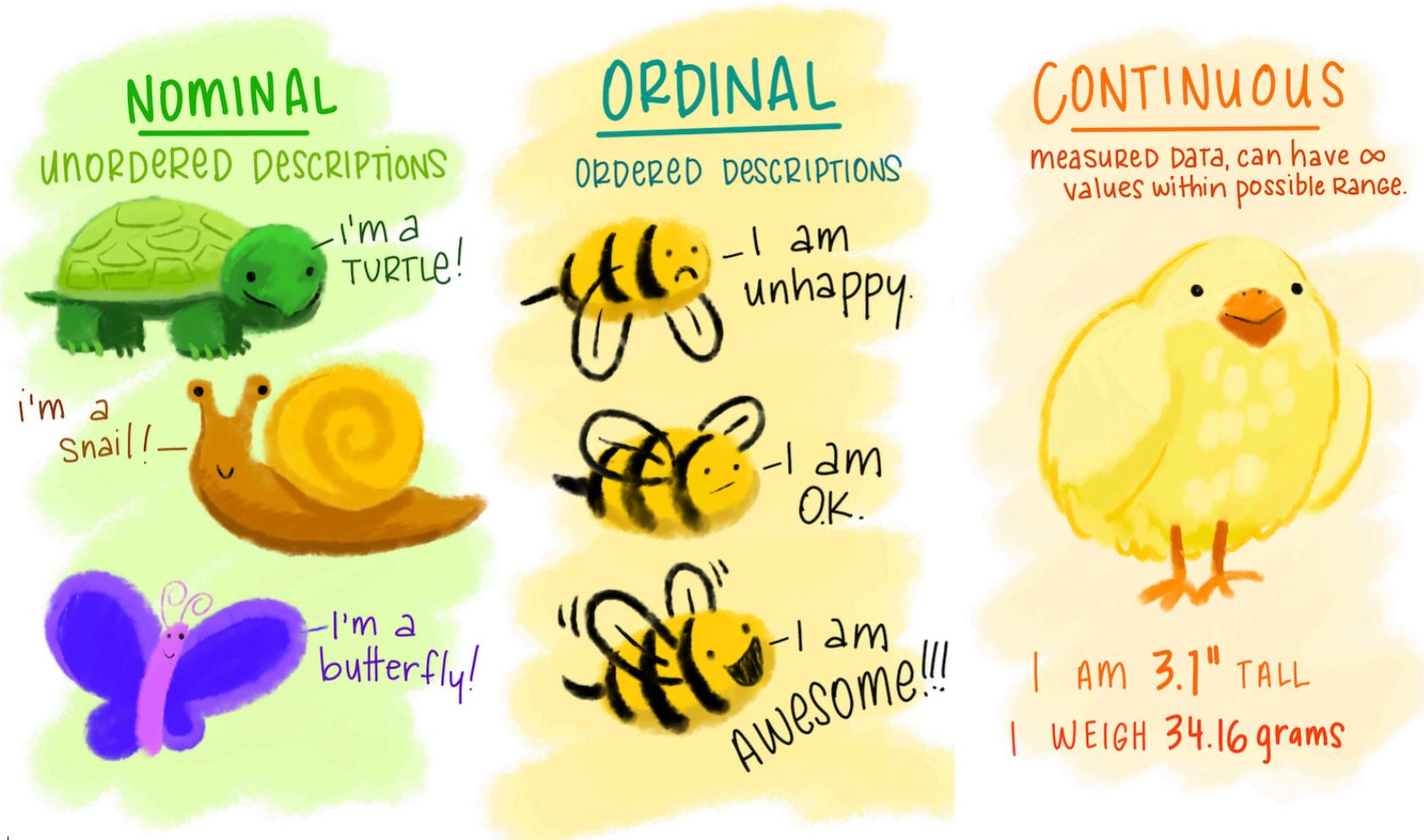
Showing 1 to 15 of 144 entries, 22 total columns

Each **cell** is a **value**, e.g. the duration of the song *for this observation*

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

Showing 1 to 15 of 144 entries, 22 total columns

Variables (columns) can be put into three different categories: categorical (nominal), ordinal, and continuous



Art by Allison Horst

Examples of categorical variables might be track name, whether it is explicit, or album name

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

Showing 1 to 15 of 144 entries, 22 total columns

Examples of **ordinal** variables might be time signature or track number

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

Showing 1 to 15 of 144 entries, 22 total columns

Examples of **continuous** variables might be duration or tempo

tempo	time_signature	duration_ms	explicit	track_name	track_number	album_name
76.009	4	232106	FALSE	Tim McGraw	1	Taylor Swift
105.586	4	173066	FALSE	Picture To Burn	2	Taylor Swift
99.953	4	203040	FALSE	Teardrops On My Guitar – Radio Single Remix	3	Taylor Swift
115.028	4	199200	FALSE	A Place in this World	4	Taylor Swift
175.558	4	239013	FALSE	Cold As You	5	Taylor Swift
112.982	4	207106	FALSE	The Outside	6	Taylor Swift
146.165	4	248106	FALSE	Tied Together with a Smile	7	Taylor Swift
131.597	4	236053	FALSE	Stay Beautiful	8	Taylor Swift
167.964	4	242200	FALSE	Should've Said No	9	Taylor Swift
74.900	4	213080	FALSE	Mary's Song (Oh My My My)	10	Taylor Swift
89.011	4	201106	FALSE	Our Song	11	Taylor Swift
100.019	4	242826	FALSE	Fearless	1	Fearless
95.431	4	294880	FALSE	Fifteen	2	Fearless
119.054	4	234213	FALSE	Love Story	3	Fearless
116.001	4	254826	FALSE	Hev Stephen	4	Fearless

Showing 1 to 15 of 144 entries, 22 total columns

If you're struggling to recognise the difference, here's something that helps:



Runner 3

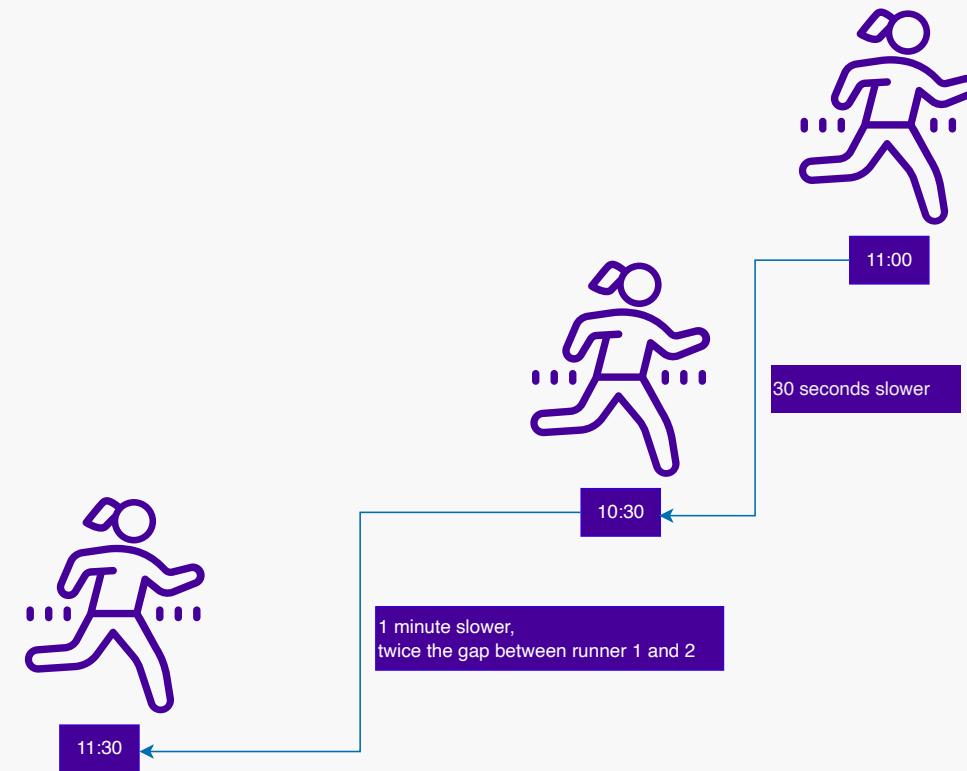


Runner 2

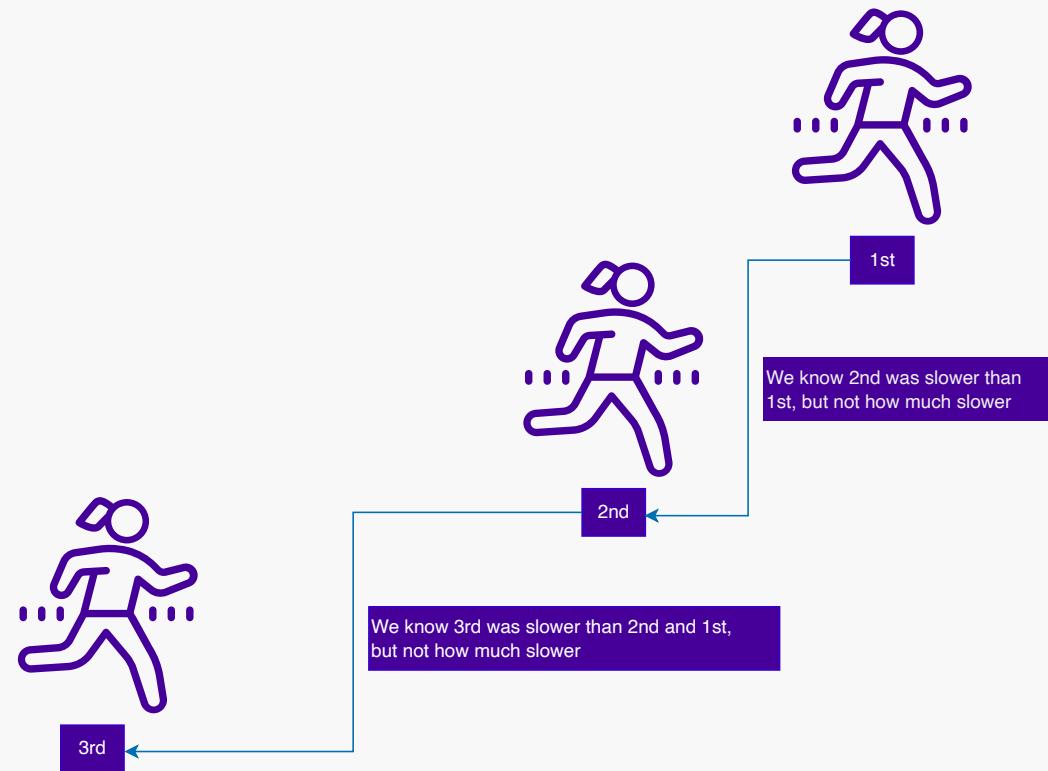


Runner 1

One variable we might have for runners is finish time: this is continuous because we know that if Runner 2 was 30 seconds slower than Runner 1, and Runner 3 was 1 minute slower than Runner 2, Runner 3 was behind by twice as much. 1 second means the same thing.



Another variable might be the placement of the runners: 1st, 2nd, 3rd. We know that runner 2 was slower than runner 1, but we don't know *how much by*. 1st, 2nd, and 3rd also means different things depending on the *race*. 3rd in a Park Run \neq 3rd in Olympics. This is ordinal.



Lastly, another variable we could have is the country the runner is representing. There is no inherent ordering to countries, any ordering would depend on some other external variable (such as alphabetical order).



There is no inherent ordering to the categories,
neither is 'more' or 'less' than the other without
some other information

Part II

The 'classic' approach: visualisation determined by the type of variables

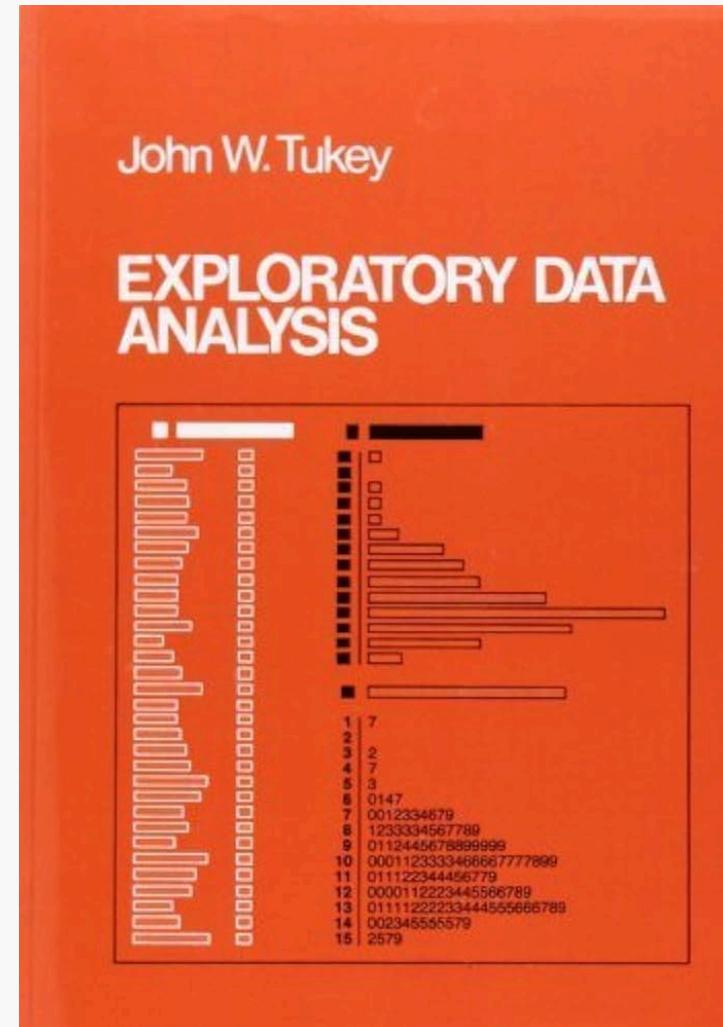
As found in many, many, introduction to quantitative research textbooks, but largely based on the work of John Tukey (1997) *Exploratory data analysis*



Classic approaches to data visualisation

These are data visualisations that are often driven by:

- The types of the variables being visualised (how many, categorical, continuous, ordinal).
- The computing power available at the time or even the effort to draw the visualisation by hand.



In this approach, depending on the type of variable, we can identify an appropriate chart type:

One variable

Variable Type	Visualisation
Nominal	Bar Chart
Ordinal	Bar Chart
Continuous	Histogram

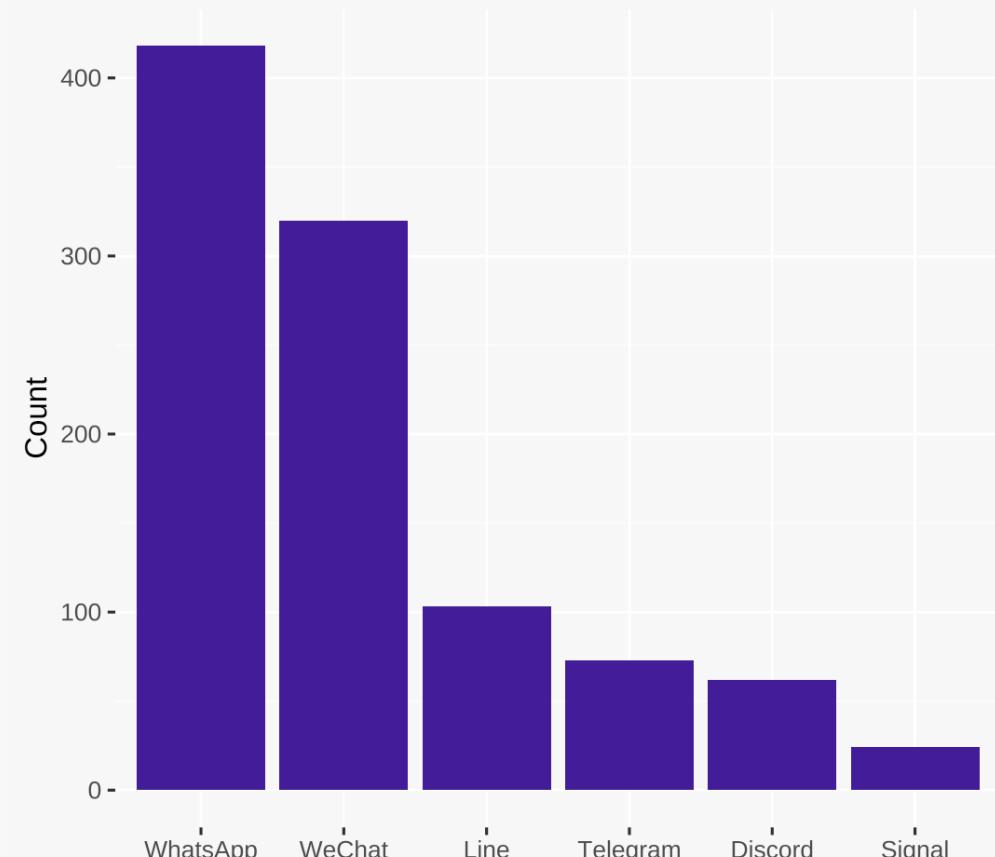
Two variables

Variable Type	Nominal	Ordinal	Continuous
Nominal	Bivariate Bar Chart		
Ordinal	Bivariate Bar Chart	Bivariate Bar Chart	
Continuous	Boxplot	Boxplot	Scatterplot

Some examples:

- Favourite messaging platform (categorical)

What are students' favour messaging platforms?

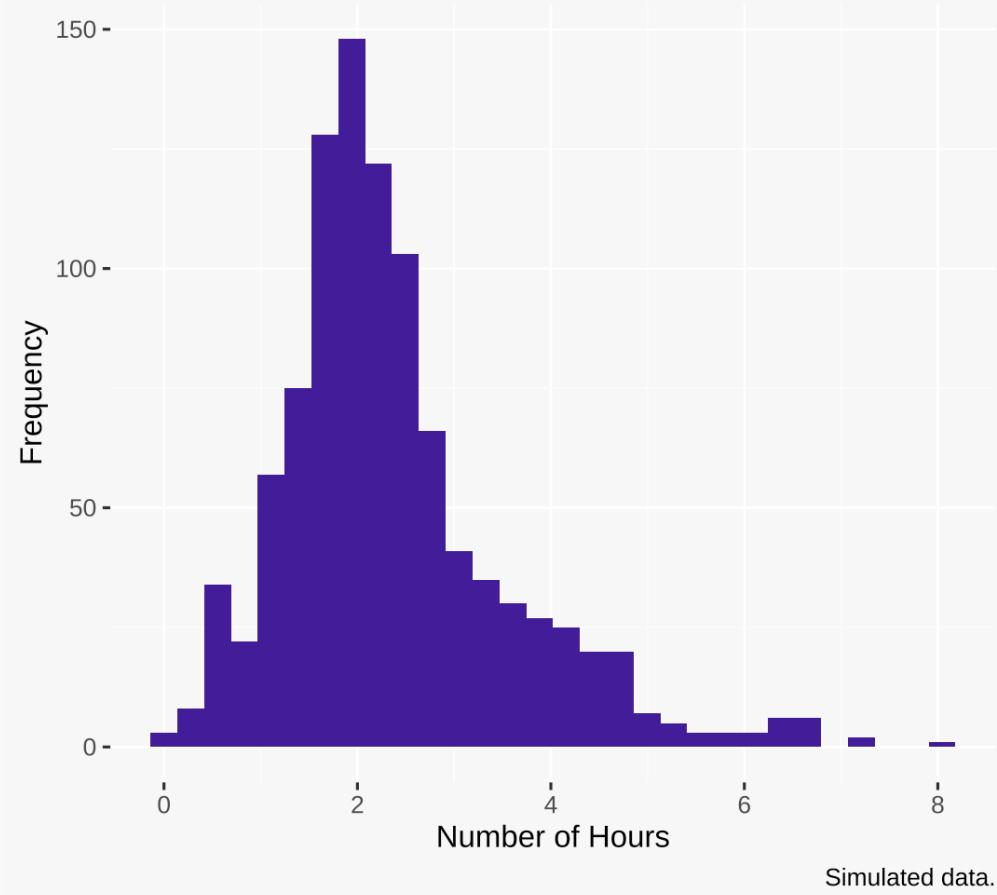


Simulated data.

Some examples:

- Favourite messaging platform (categorical)
- Number of hours spent on social media per day (continuous)

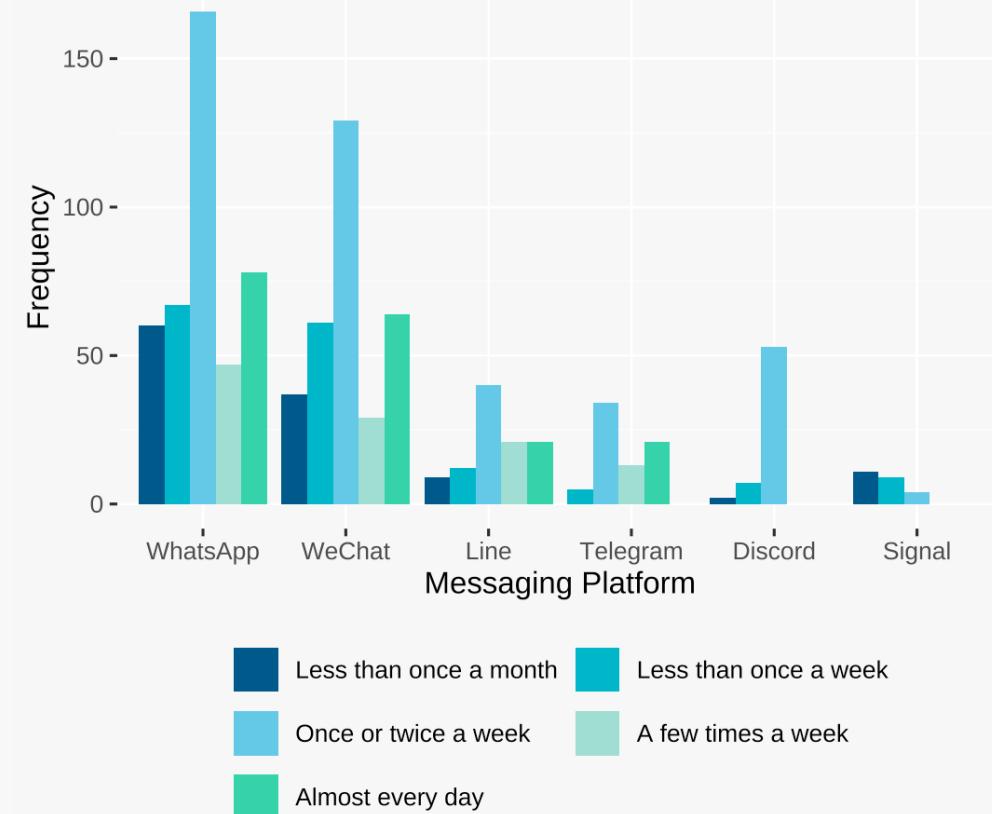
How many hours a day are spent on messaging platforms?



Some examples:

- Favourite messaging platform (categorical)
- Number of hours spent on social media per day (continuous)
- Do students who use some instant messaging platforms meet up in person less often than others? (categorical × ordinal)

How often do students using different messaging platforms meet up with friends they don't live with?

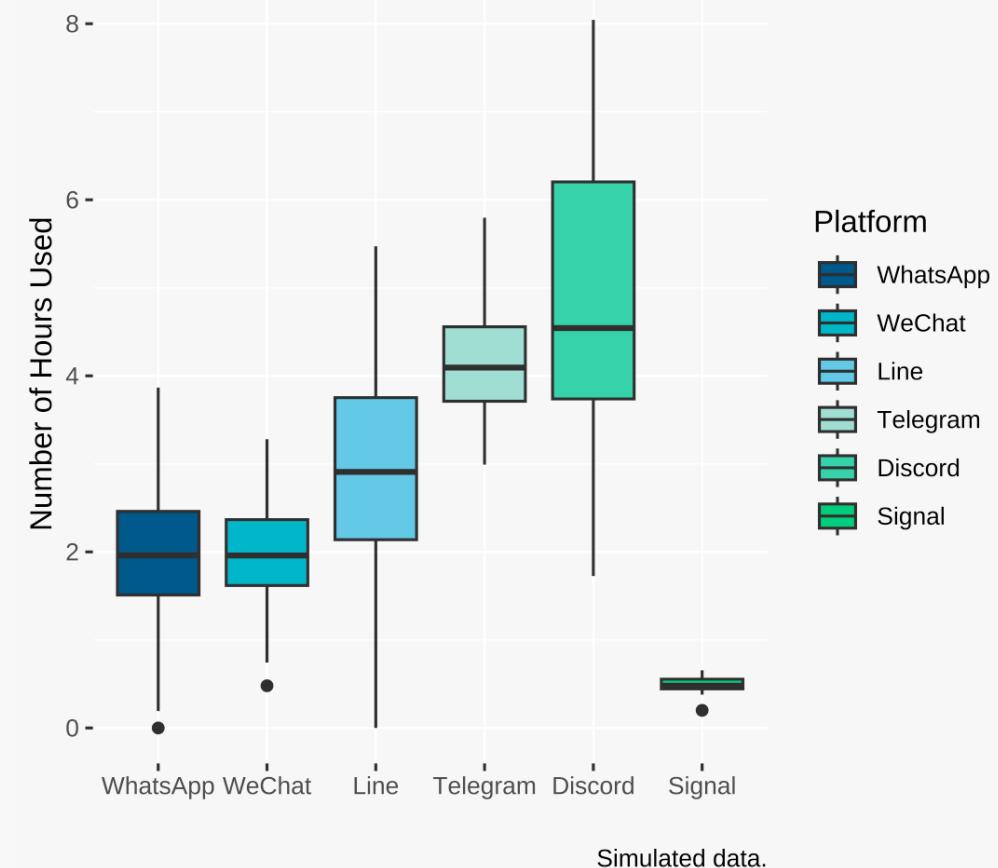


Simulated data.

Some examples:

- Favourite messaging platform (categorical)
- Number of hours spent on social media per day (continuous)
- Do students who use some instant messaging platforms meet up in person less often than others? (categorical \times ordinal)
- Do students spend different amounts of time on instant messaging platforms depending on the platform they use? (categorical \times continuous)

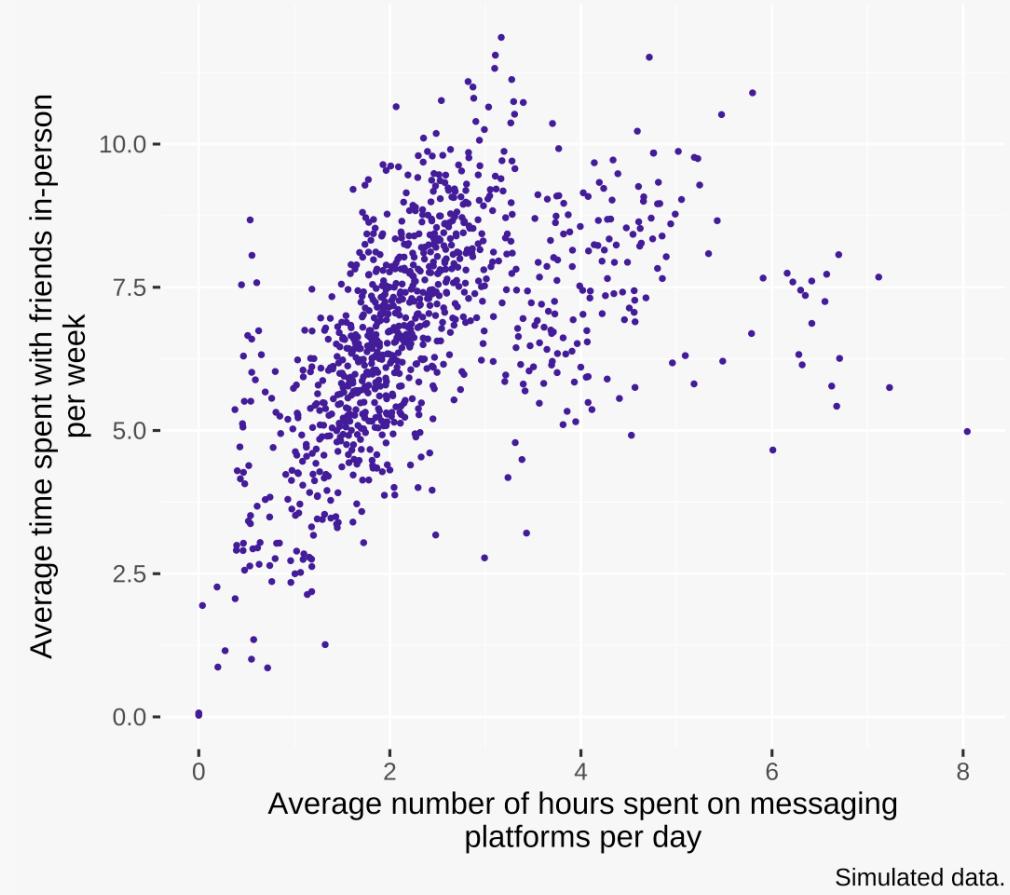
Is the choice of messaging platform linked to time spent using it?



Some examples:

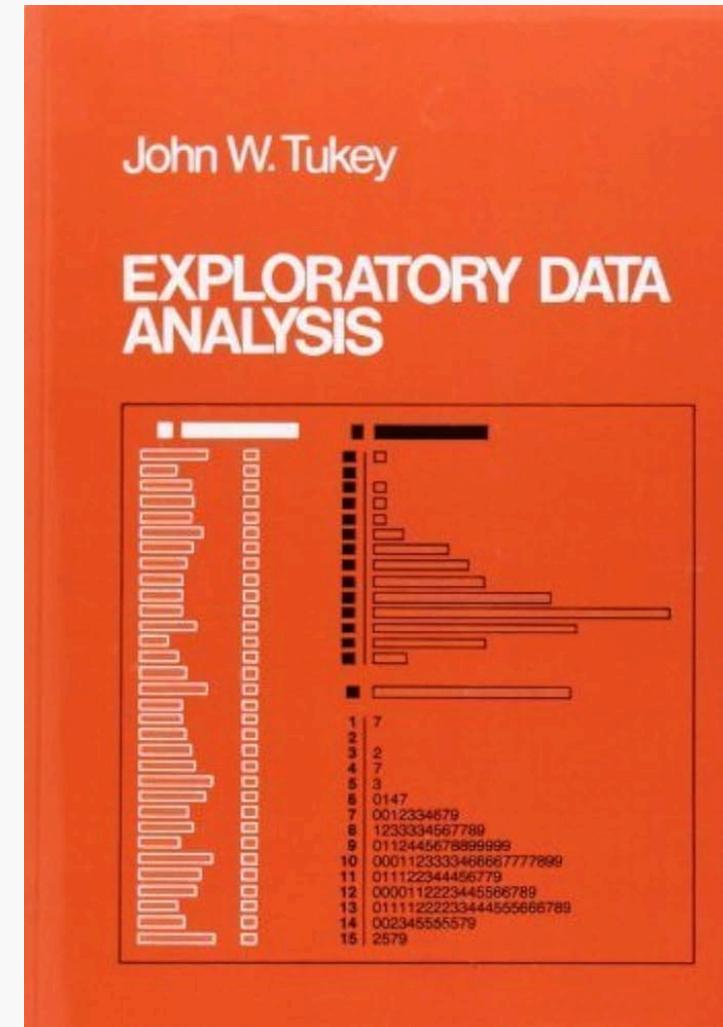
- Favourite messaging platform (categorical)
- Number of hours spent on social media per day (continuous)
- Do students who use some instant messaging platforms meet up in person less often than others? (categorical \times ordinal)
- Do students spend different amounts of time on instant messaging platforms depending on the platform they use? (categorical \times continuous)
- Do students who spend more time on instant messaging platforms spend less time in-person with people they don't live with? (continuous \times continuous)

Do students who spend more time on messaging platforms spend less time with friends in person?



Classic approaches to data visualisation

- Easy to apply
- Easy to remember
- Quite restrictive, doesn't provide multiple options
- Not always very appealing aesthetically
- Some redundancy in visual presentation
- Sometimes doesn't work so well with very large datasets (e.g. scatterplots)



Part III

Story-driven choice of chart type

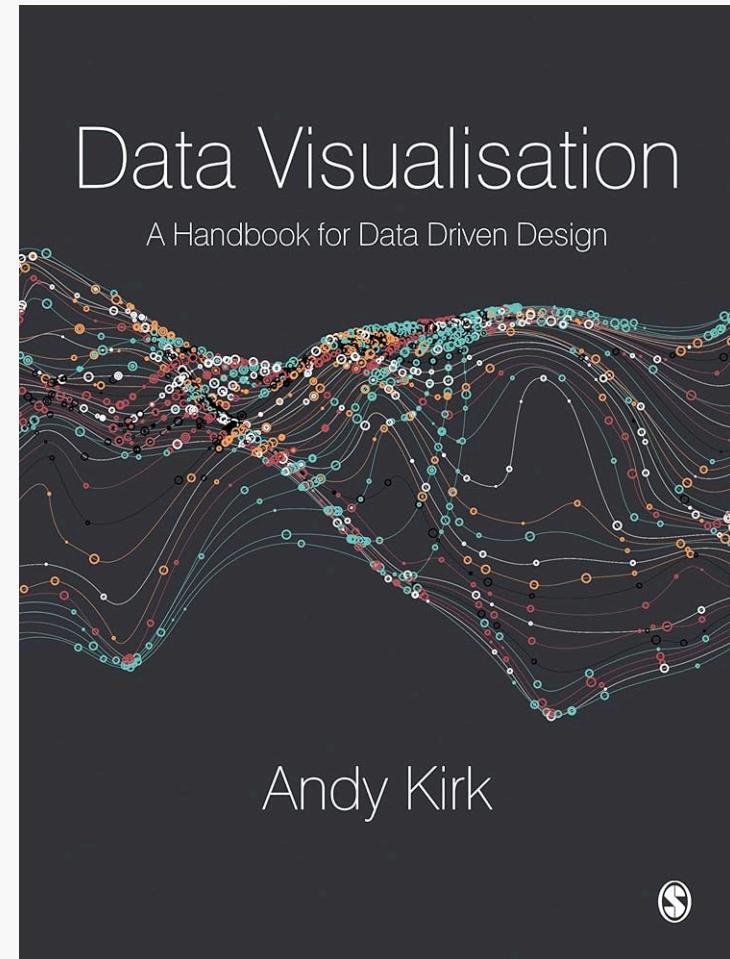
Using Andy Kirk's (2019) Data Visualisation A Handbook for Data Driven Design & FT's Visual Vocabulary (Smith, et al. 2019)



Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- **Categorical:** Comparing categories and distributions of quantitative values.
- **Hierarchical:** Revealing part-to-whole relationships and hierarchies.
- **Relational:** Exploring correlations and connections.
- **Temporal:** Plotting trends and intervals over time.
- **Spatial:** Mapping spatial patterns through overlays and distortions.



Kirk (2019)

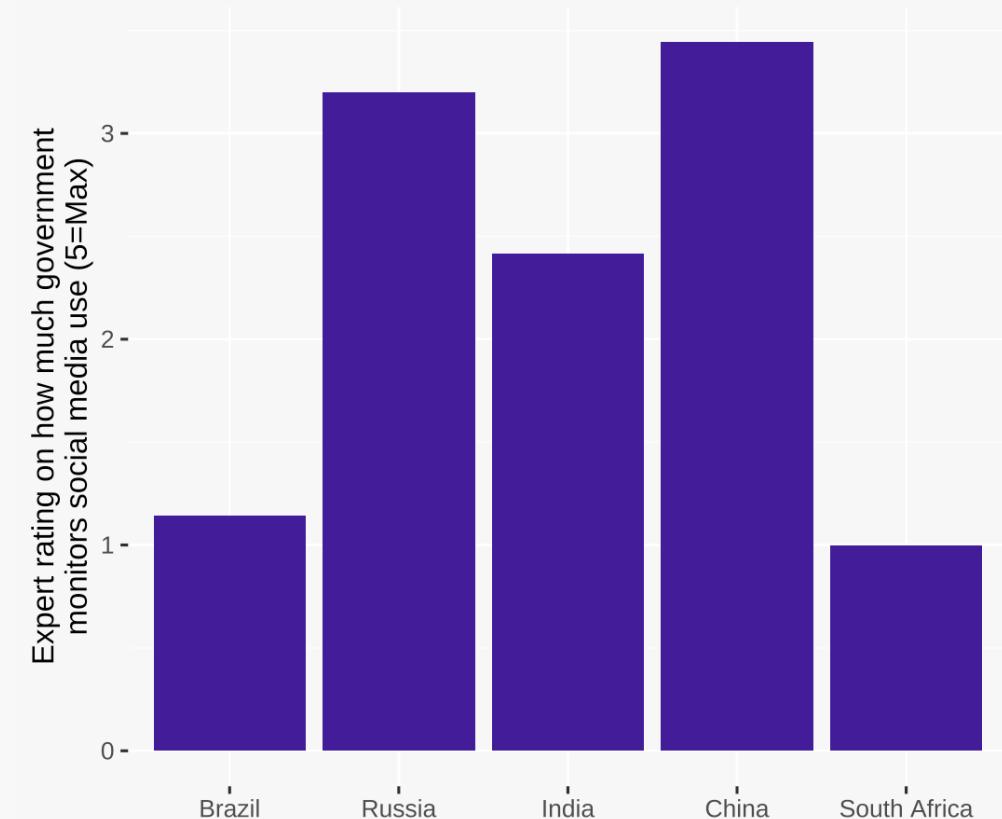
Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- **Categorical:** Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

bar chart: `geom_col` / `geom_bar`

Expert assessments of social media monitoring vary across the BRICS



Source: Mechkova, et al. (2024) Digital Society Project.

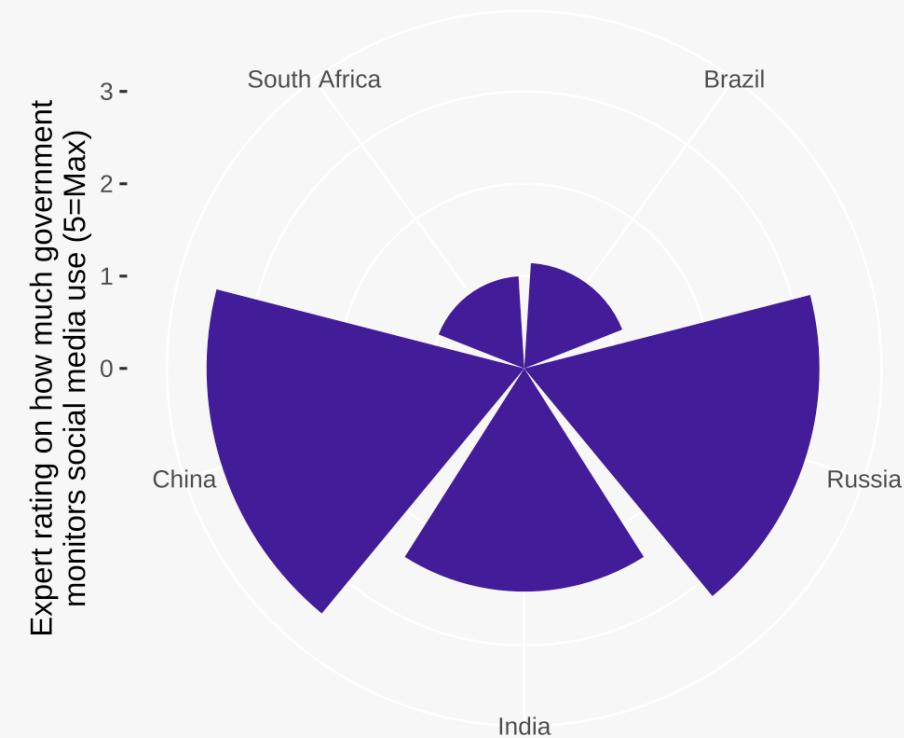
Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- **Categorical:** Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

sunburst chart: `geom_col / geom_bar + coord_polar`

Expert assessments of social media monitoring vary across the BRICS



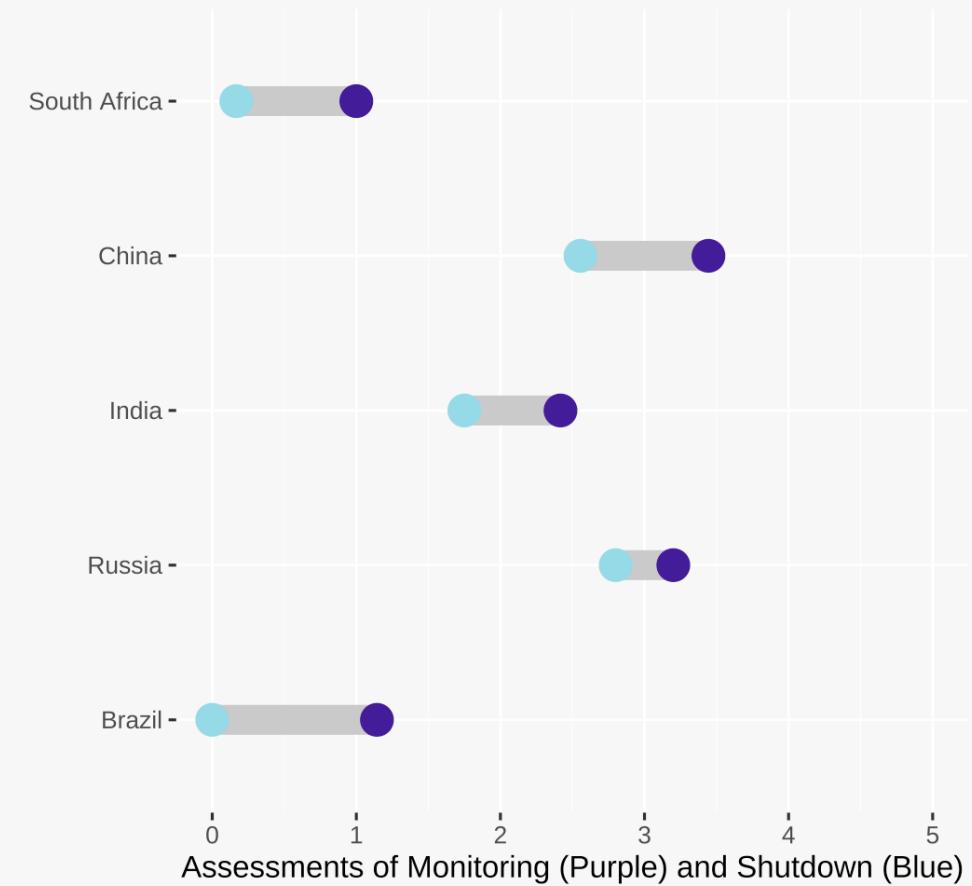
Source: Mechkova, et al. (2024) Digital Society Project.

Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- **Categorical:** Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

dumbbell chart: `geom_point` x2 + `geom_segment`

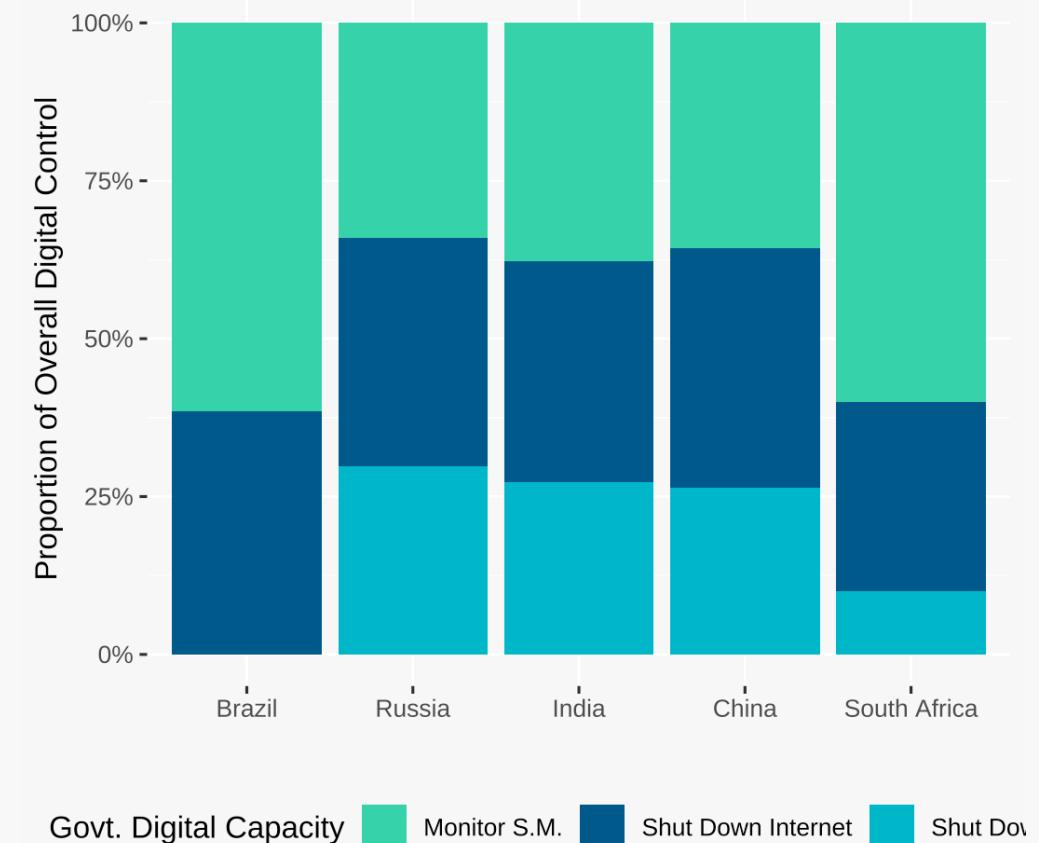


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- **Hierarchical:** Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

100% Stacked Bar Chart: `geom_col/bar + position = "fill"`

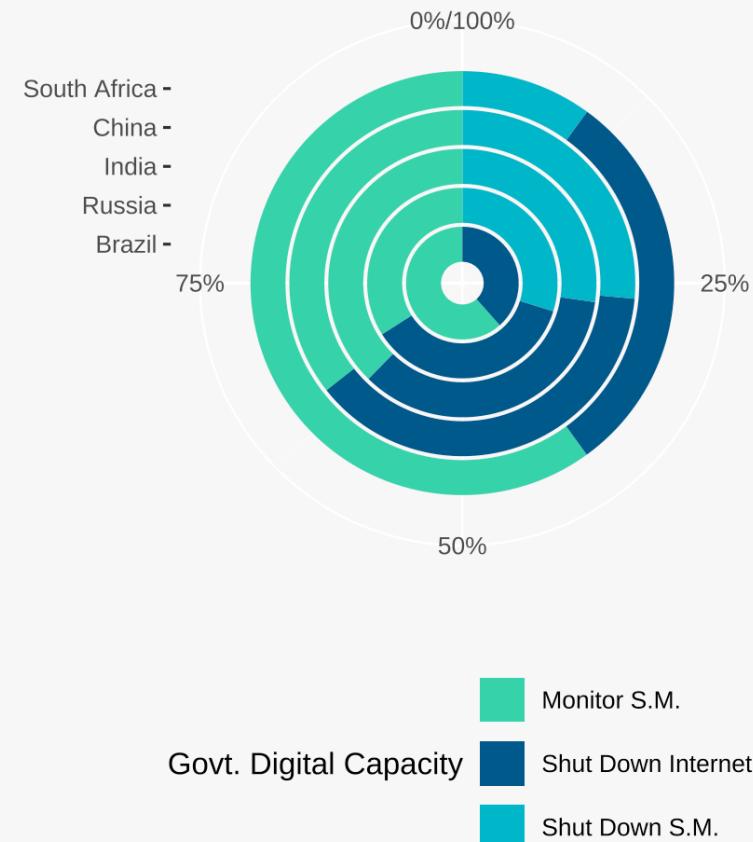


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- **Hierarchical:** Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Donut Chart: Horizontal 100% Stacked Bar `geom_col/bar` + `coord_polar`



Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- **Hierarchical:** Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Dendrogram using `ggdendro`

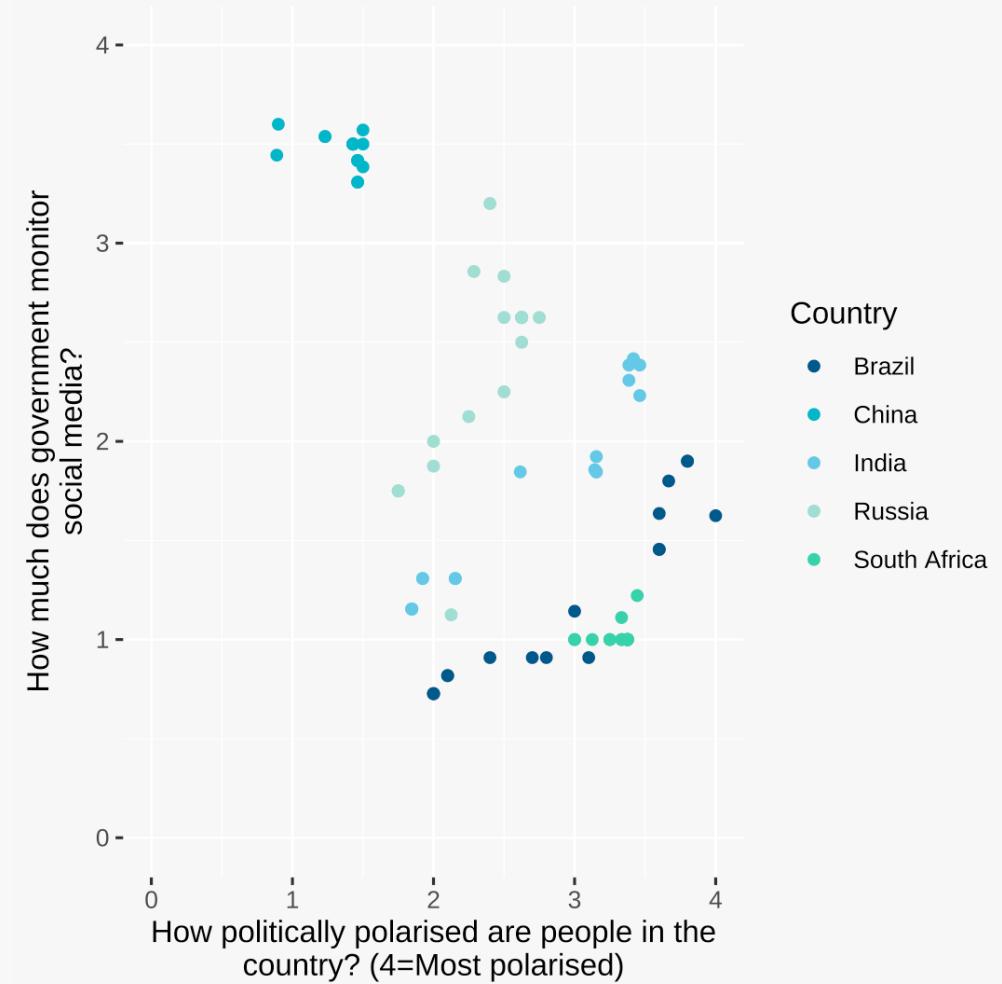


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- **Relational:** Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Scatterplot: `geom_point`

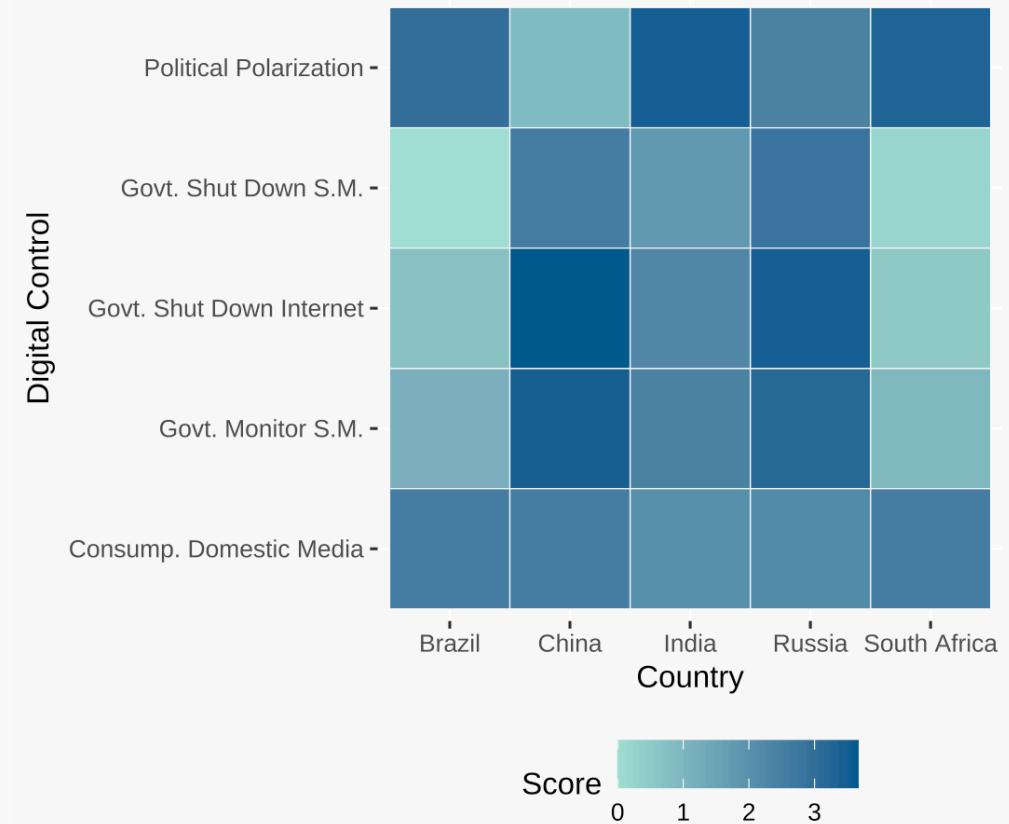


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- **Relational:** Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Heatmap: `geom_tile`

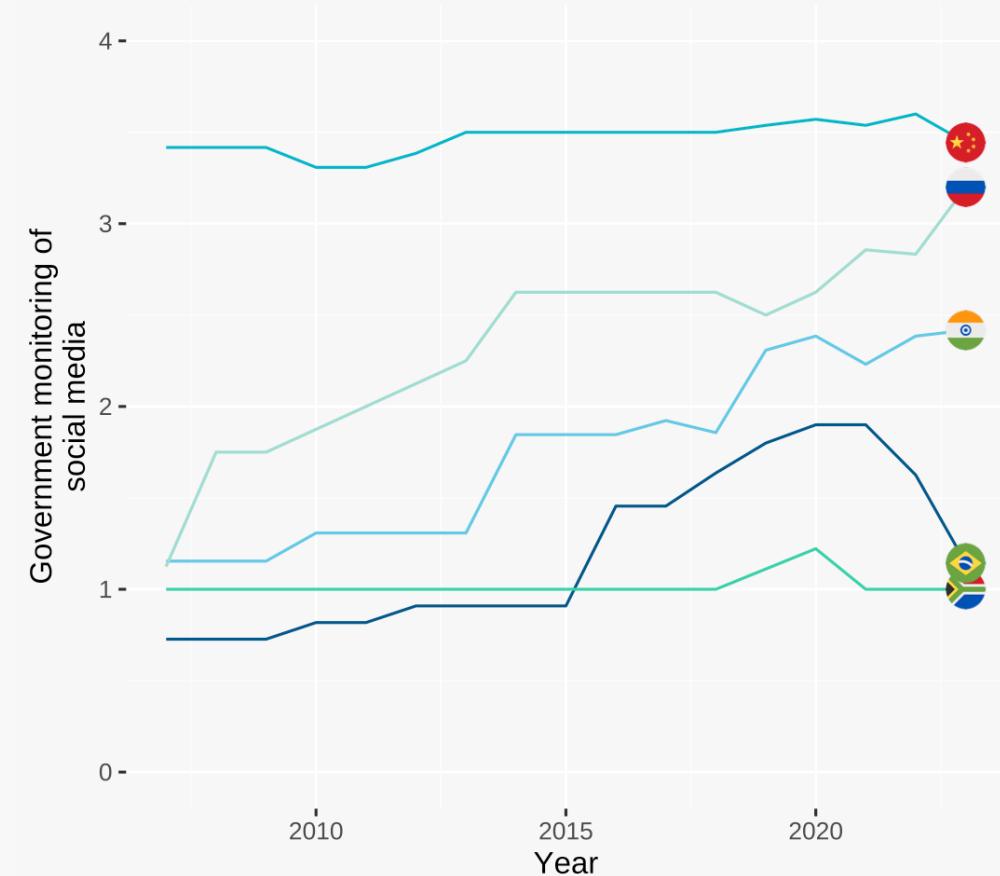


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- **Temporal**: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Line plot: `geom_line + group / colour aes`

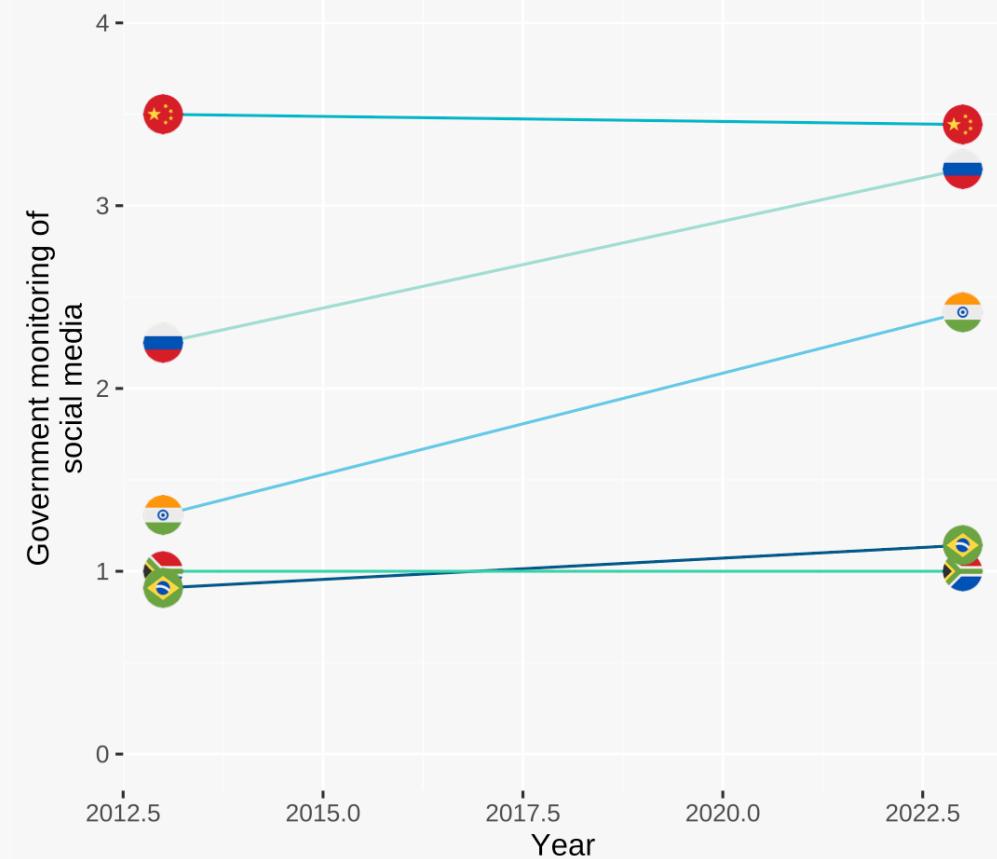


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- **Temporal**: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Line plot: `geom_line + group / colour aes`

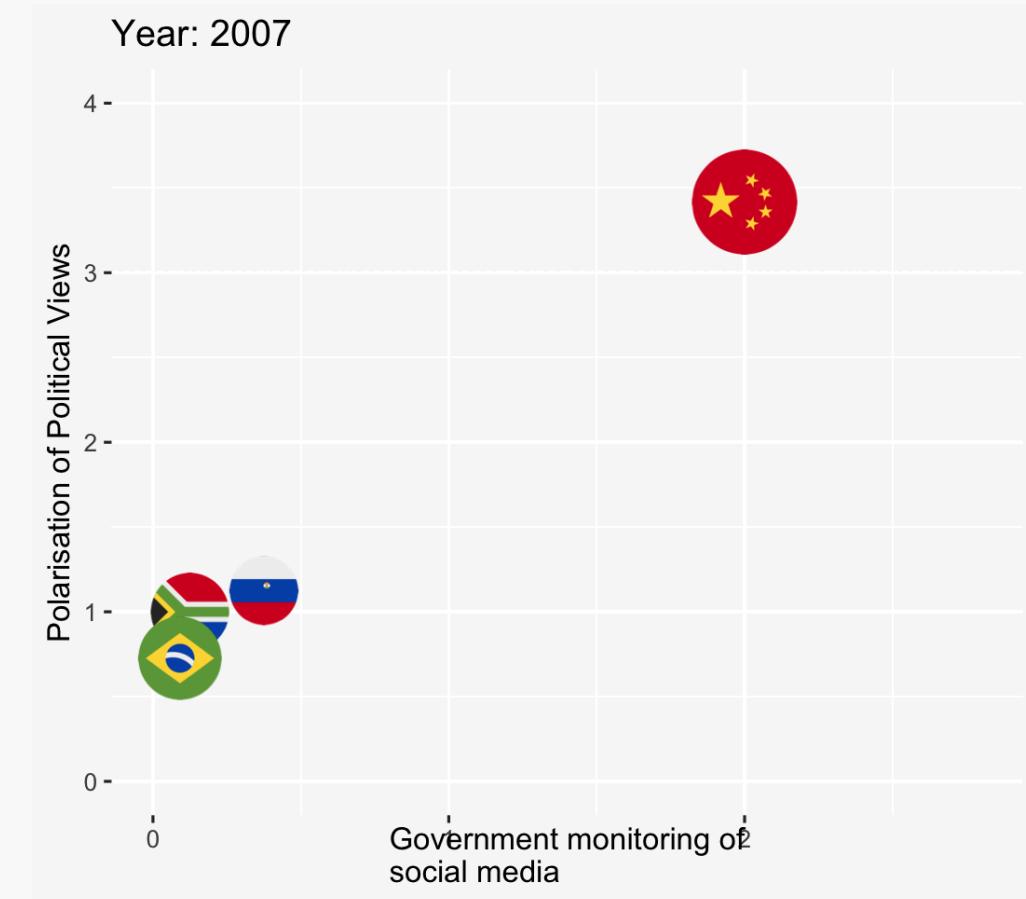


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- **Temporal**: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Moving plot: `gganimate`

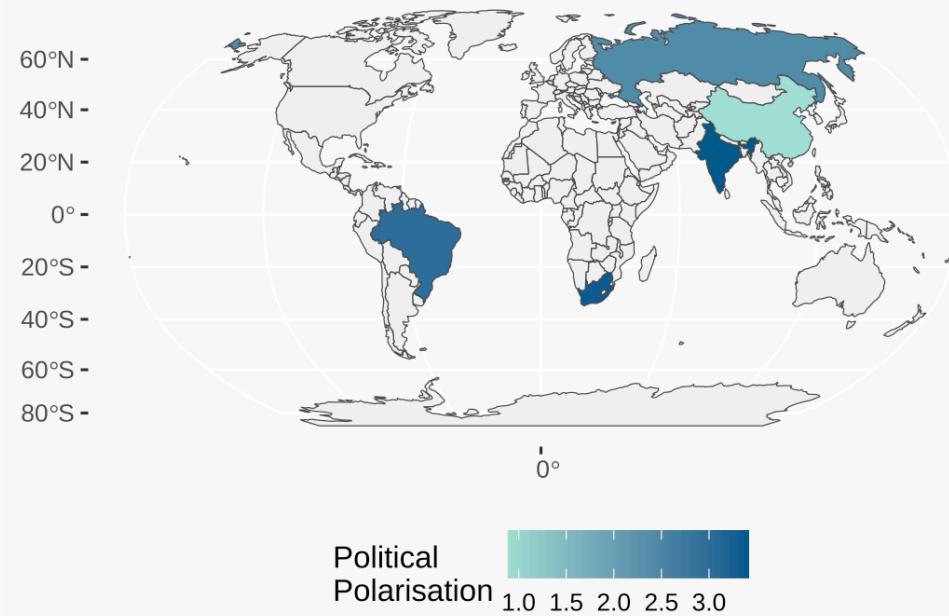


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- **Spatial:** Mapping spatial patterns through overlays and distortions.

Choropleth: `geom_sf`

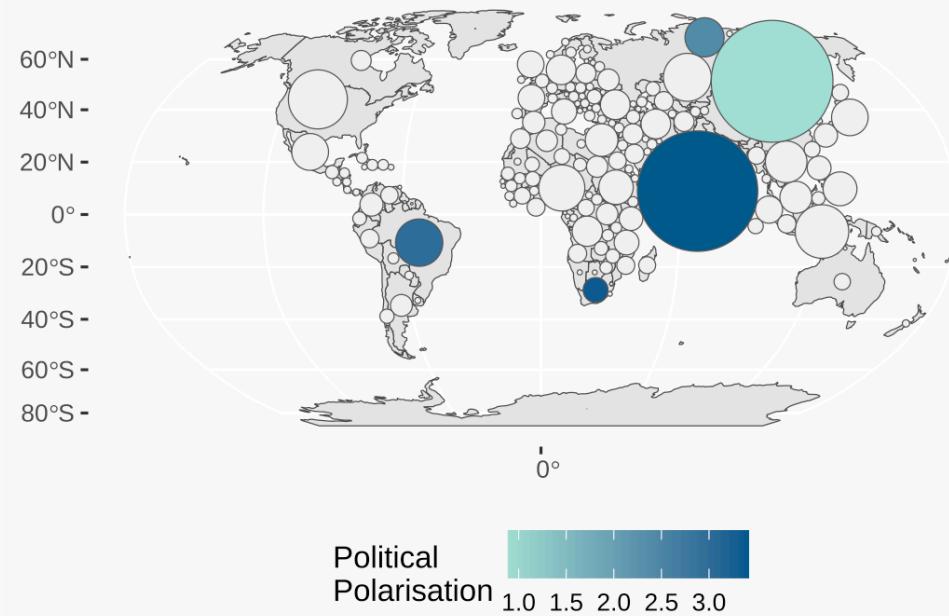


Andy Kirk's CHRTS

While the 'data-driven' approach to selecting a visualisation works well, we now have many, many, different types of visualisation we could choose for the same types of data. It can make more sense to start thinking about what kind of story we want to tell.

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- **Spatial:** Mapping spatial patterns through overlays and distortions.

Dorling Cartogram: `cartogram` package & `geom_sf`



We can use both the types of data that we have and the type of story we are trying to tell to identify appropriate ways to visualise our data.

The rest of this week:

This week's workshop:

Let's extend graphs where we show the relationship between a continuous and a categorical variable: box plots, density curves, histograms, and so on. What's the easiest way to understand this information? Why would we use one of these graphs over another?

The rest of this week:

This week's workshop:

Let's extend graphs where we show the relationship between a continuous and a categorical variable: box plots, density curves, histograms, and so on. What's the easiest way to understand this information? Why would we use one of these graphs over another?

Before week 5:

Core tasks:

- **Before lecture, read at least one of these:**
 - On the basics of formatting charts:
<https://analysisfunction.civilservice.gov.uk/policy-store/data-visualisation-charts/#section-6> (ignore the last section 'Communicating quality and uncertainty in charts')
 - On working with text: <https://blog.datawrapper.de/text-in-data-visualizations/>
 - On working with colour:
<https://blog.datawrapper.de/emphasize-with-color-in-data-visualizations/>

Before workshop: Read the assessment brief for assessment 1

Before workshop: Work through the task at the end of the workshop handout

Supplementary tasks: Before workshop: Read chapter 4/Show the right numbers of the Healy book

The rest of this week:

This week's workshop:

Let's extend graphs where we show the relationship between a continuous and a categorical variable: box plots, density curves, histograms, and so on. What's the easiest way to understand this information? Why would we use one of these graphs over another?

Before week 5:

Core tasks:

- **Before lecture, read at least one of these:**
 - On the basics of formatting charts:
<https://analysisfunction.civilservice.gov.uk/policy-store/data-visualisation-charts/#section-6> (ignore the last section 'Communicating quality and uncertainty in charts')
 - On working with text: <https://blog.datawrapper.de/text-in-data-visualizations/>
 - On working with colour:
<https://blog.datawrapper.de/emphasize-with-color-in-data-visualizations/>

Before workshop: Read the assessment brief for assessment 1

Before workshop: Work through the task at the end of the workshop handout

Supplementary tasks: Before workshop: Read chapter 4/Show the right numbers of the Healy book