Efficiency of Stream Graphs

By Holly Echlin and Samantha Roberts
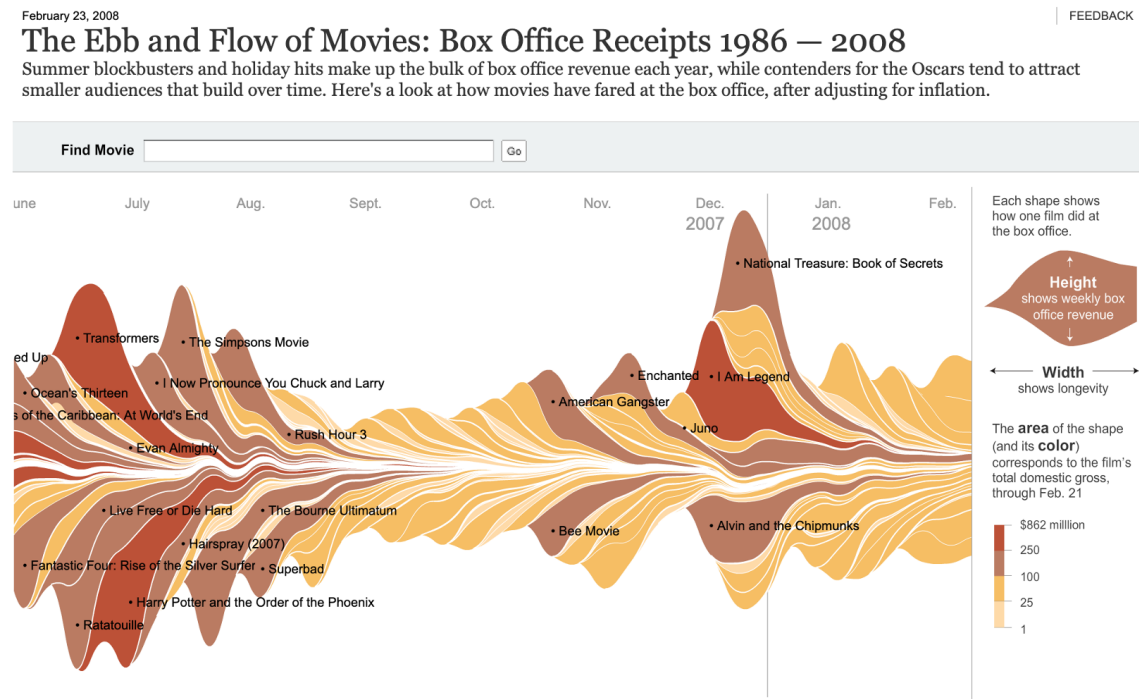
PSYC 6135

Stream Graphs

- Popularized in 2008 by *New York Times* journalist in an article displaying box office revenues (Bloch, Byron, Carter, & Cox, 2008)

- Variation of the stacked area chart

- Organic shapes - "streams" - displaying change over time

- Visually appealing

NYT Journalist: Amanda Cox

February 23, 2008   FEEDBACK

The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.

Benefits of Using Stream Graphs

- Minimization of 'wiggle' in data points seen in the stacked bar graph
- Provides the ability to look at data of multiple groups overtime
- Shows all data points rather than magnitude data at specific time points
- More visually appealing and increased readability (Thudt et al., 2016)

Focus of Covid-19 deaths has switched from Asia to Europe – and now the US

Daily deaths of patients diagnosed with coronavirus

April 14
Total daily deaths
5,100

The **US** now accounts for nearly **one-third** of global daily deaths

Apr 14
1,541

US

March 1
Total daily deaths
58

Rest of N America

Spain

Italy

UK

Rest of Europe

The **UK** now records more daily deaths than either Spain or Italy

LatAm and Caribbean

Asia

Africa

Mid East

Mar 1   Mar 8   Mar 15   Mar 22   Mar 29   Apr 5   Apr 14

Drawbacks to Stream Graphs

- Easily cluttered
- Does not support negative values (or combinations of negative and positive values together)
  - No clear baseline to differentiate positive and negative data values
- Challenging to read the scale at a glance due to the lack of y-axis
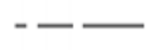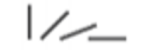- Difficult to draw relationships between variables

Stacked Bar Graph

Stream Graph

Equivocal Literature Comparing Stacked Bar and Stream Graphs

- Stream graphs have been found to smooth distortions of each stream which has been found to improve the readability over ThemeRiver but not stacked bar graphs (Thudt, 2016)
- Stream graphs have been found to reduce the time-series distortion which has been found to improve the readability over basic stacked bar graphs (Byron & Wattenberg, 2008; VanderPlaus, 2015)
- Analyzed several examples of causal stream graphs published on the web and concluded that stream graphs are difficult to read due to their uncommon shapes (Kirk, 2010)
- Stream graphs have been found to be more aesthetically appealing; however, the low-distortion of stream graphs based on the perception of curved slopes and slope ratios have been found not to improve the readability over stacked bar graphs (Bartolomeo & Hu, 2016)

Perceptual Factors

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

Fig. 5.6 from: Munzner, *Visualization Analysis & Design*

24

POSITION COMMON SCALE | POSITION NON-ALIGNED SCALES | LENGTH

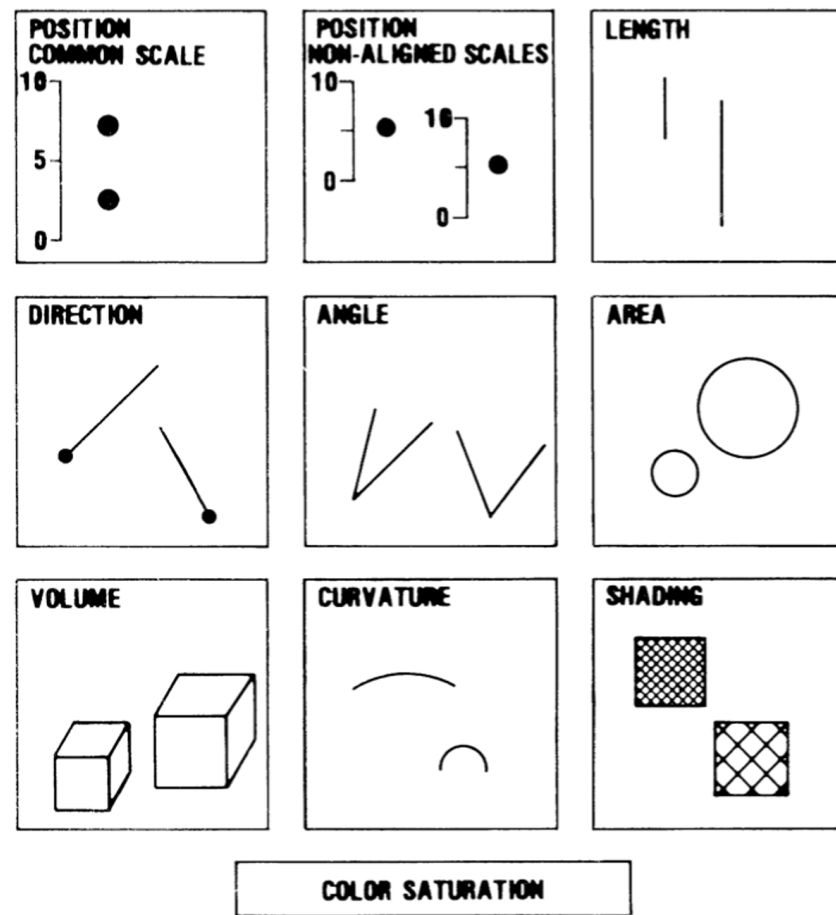DIRECTION | ANGLE | AREA

VOLUME | CURVATURE | SHADING

COLOR SATURATION

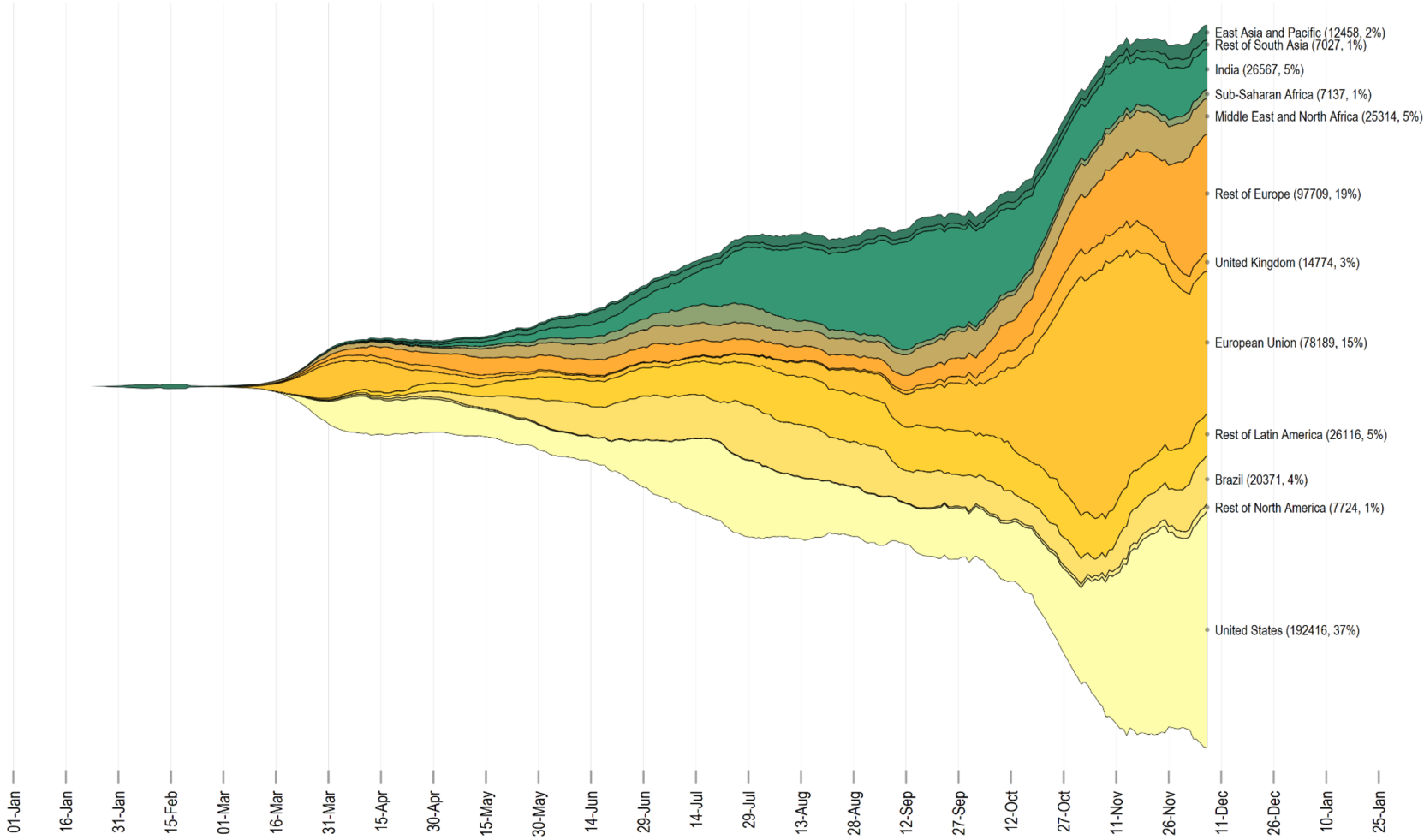Figure 1. Elementary perceptual tasks.

Stream Graphs in the Media:

COVID-19 Data

COVID-19 Daily Cases - The World

East Asia and Pacific (12458, 2%)
Rest of South Asia (7027, 1%)
India (26567, 5%)
Sub-Saharan Africa (7137, 1%)
Middle East and North Africa (25314, 5%)

Rest of Europe (97709, 19%)

United Kingdom (14774, 3%)

European Union (78189, 15%)

Rest of Latin America (26116, 5%)

Brazil (20371, 4%)
Rest of North America (7724, 1%)

United States (192416, 37%)

01-Jan 16-Jan 31-Jan 15-Feb 01-Mar 16-Mar 31-Mar 15-Apr 30-Apr 15-May 30-May 14-Jun 29-Jun 14-Jul 29-Jul 13-Aug 28-Aug 12-Sep 27-Sep 12-Oct 27-Oct 11-Nov 26-Nov 11-Dec 26-Dec 10-Jan 25-Jan
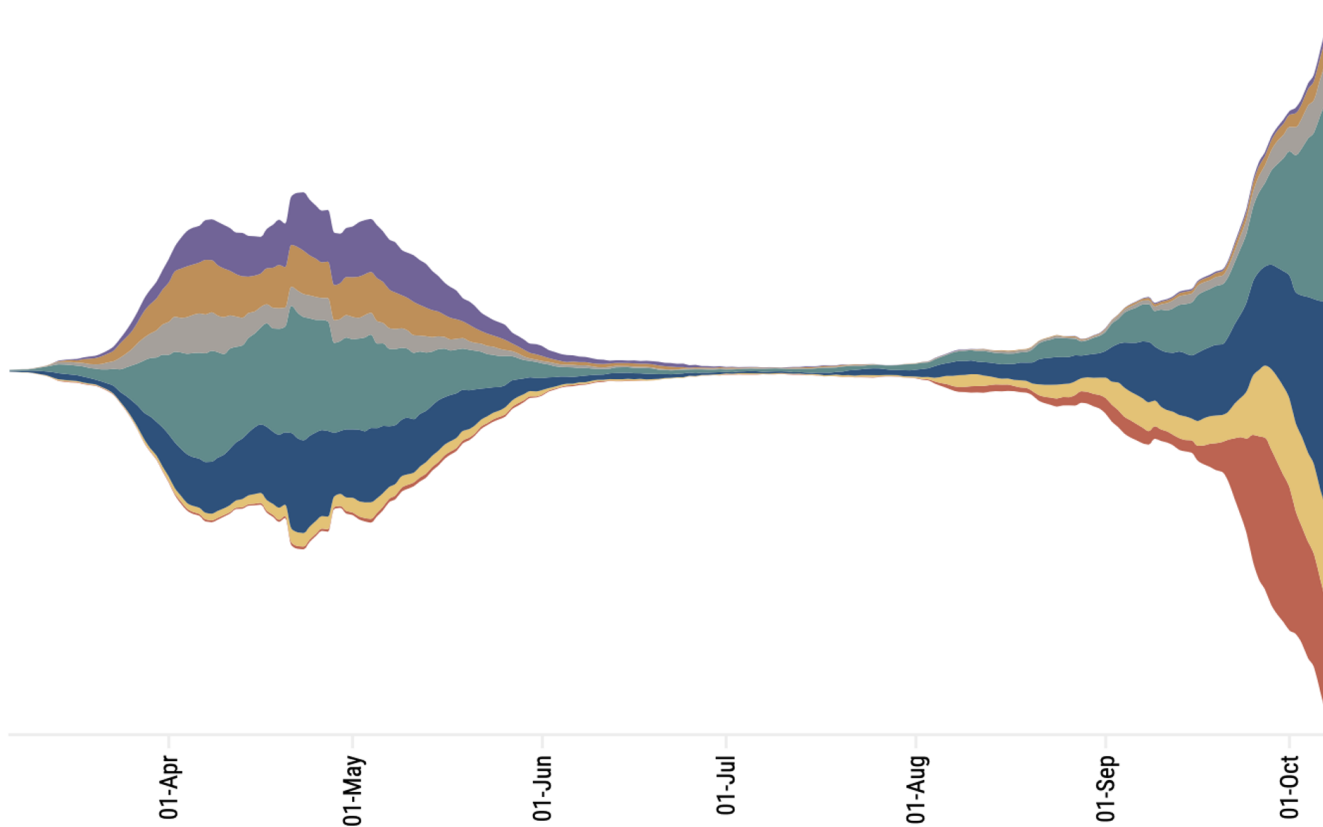
Data sources: Our World in Data. World Bank 2020 classifications used for country groups.

Cases amongst the 15 to 19 age group have surged
Graph shows seven day rolling average of positive tests by age range

Click to remove 👉 ■ 15 to 19 ■ 20 to 24 ■ 25 to 44 ■ 45 to 64 ■ 65 to 74 ■ 75 to 84 ■ 85plus

01-Apr 01-May 01-Jun 01-Jul 01-Aug 01-Sep 01-Oct
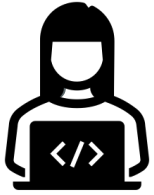
Source: Public Health Scotland • Graphic by Lesley-Anne Kelly for The Courier

Methods

Participants

- 120 adults: 60 female and 60 male
- York University Research Participant Pool
- Demographics:
  - Age
  - Sex
  - Years of post-secondary education
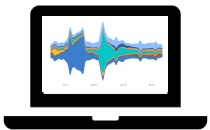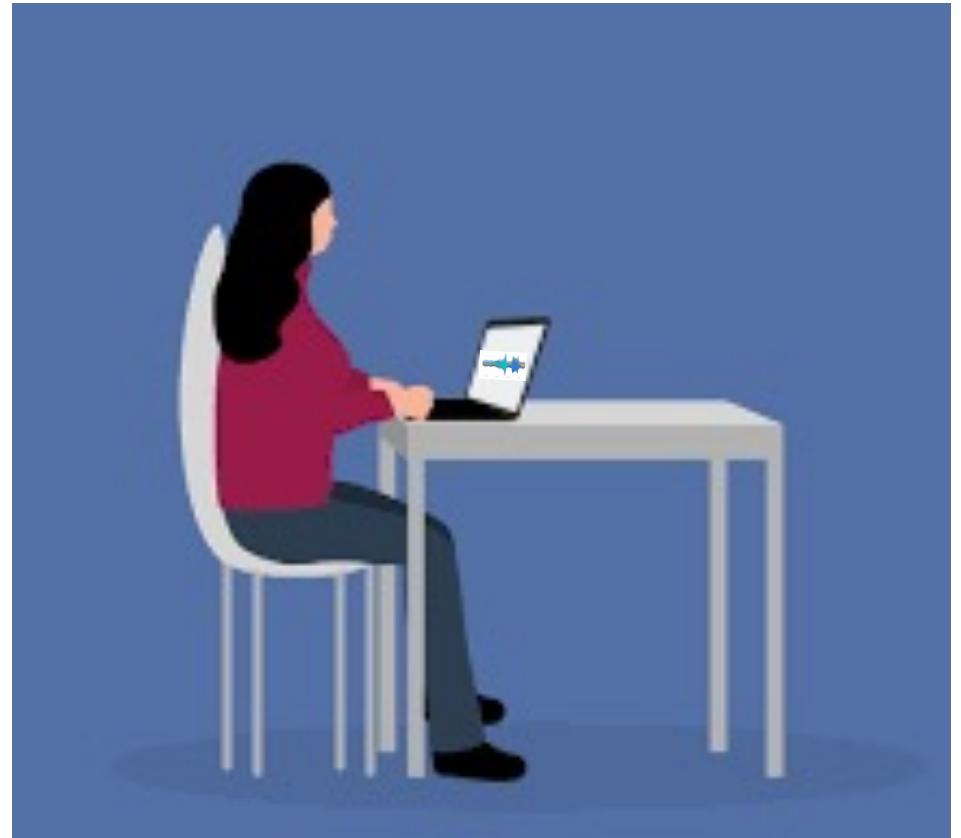 - # of statistics courses taken

Procedure

Training trial
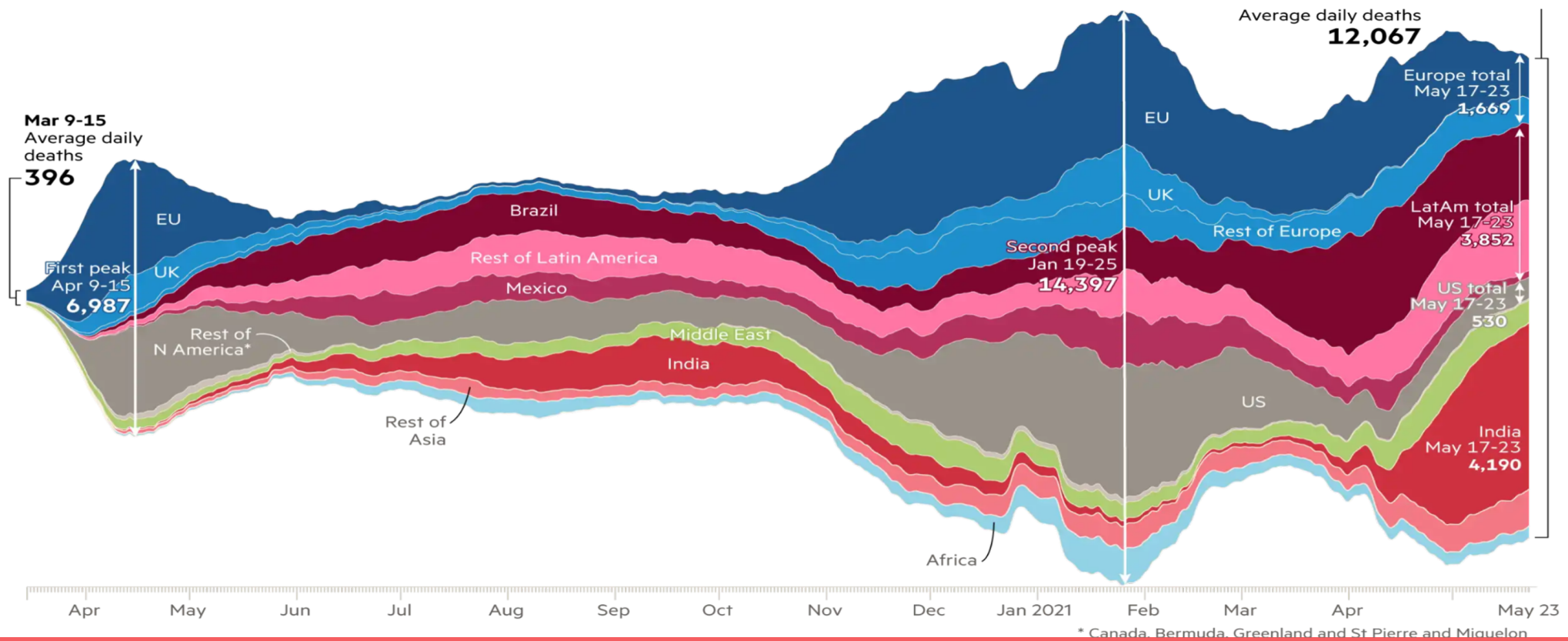
Feedback

Experimental Task

- Here is the question you must answer after viewing the graph on the upcoming screen:

At the end of January, which part of the world had the third highest number of COVID-19-positive adults?

- Please press the spacebar to continue.

- Once the upcoming graph is displayed, you will have a maximum of 30 seconds to choose your answer from a multiple-choice list on the following screen. Try to answer as quickly and accurately as possible.

- Please press the spacebar when you are ready to view the graph.

At the end of January, which part of the world had the third highest number of COVID-19-positive adults?

Participants will see everything on the same screen:

30 Seconds

Data Analyses

Average time to respond

Accuracy

- Regression analyses for each of the conditions

- Independent variables:
  - The 2 graph types, age, sex, question about graph, and previous exposure to a statistics course

Order of Counterbalancing for Both Groups

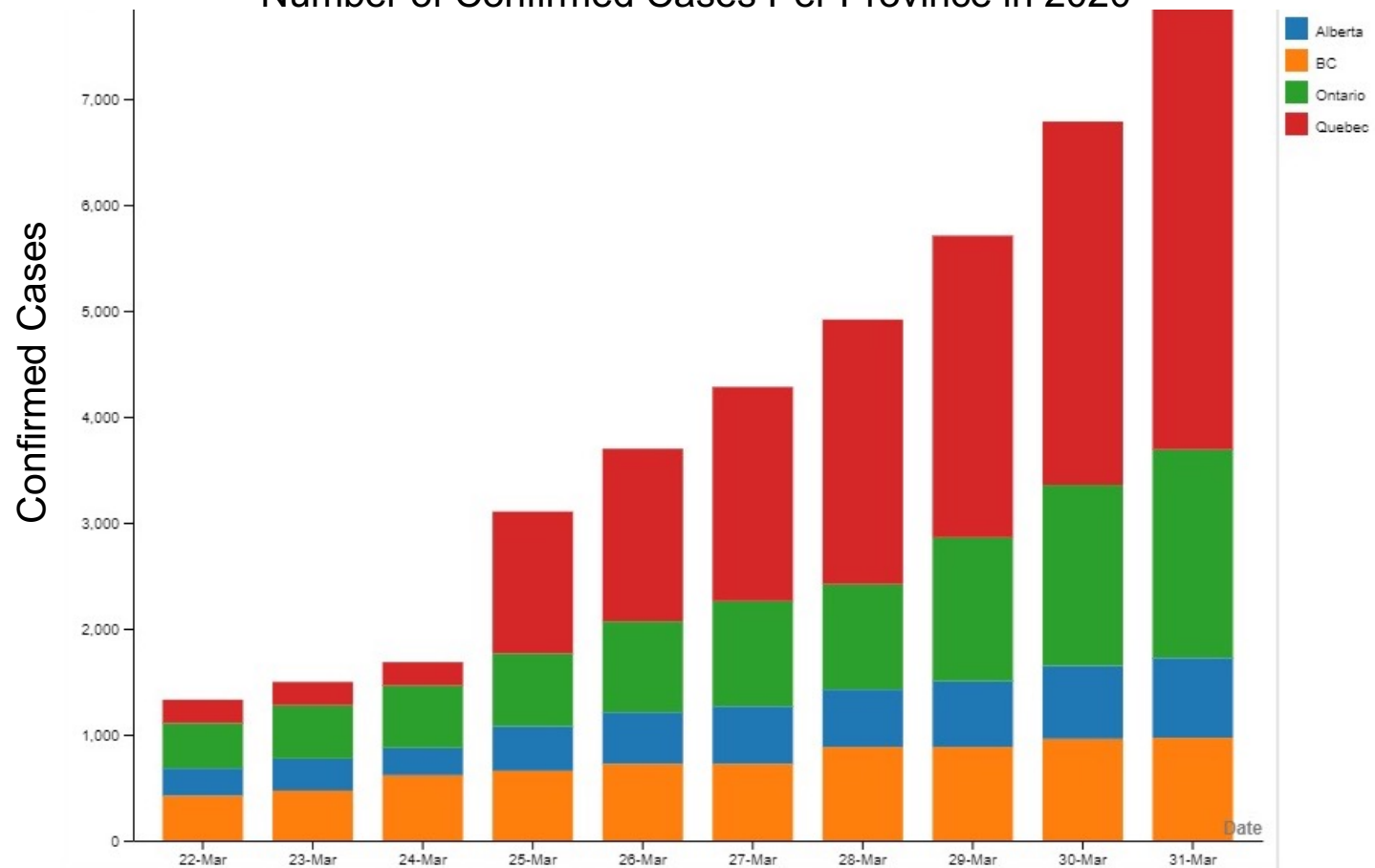|					
Group A	Stacked bar (1st question)	Stream graph (1st question)	Stacked bar (2nd question)	Stream graph (2nd question)	Stacked bar (3rd question)
Group B	Stream graph (1st question)	Stacked bar (1st question)	Stream graph (2nd question)	Stacked bar (2nd question)	Stream graph (3rd question)

Note. The order continues for 60 trials including and beyond the order shown here.

Types of Questions to Ask

- Data-related variables to assess
  - Slope (e.g., which month shows the greatest increase for BC)
  - Magnitude (e.g., on which day does Quebec have the most cases)
  - Relationship (e.g., on the 24th: what is the relationship between Quebec and Ontario)

- Graph-related variables to assess
  - Continuous versus discrete data
  - Small versus larger number of groups
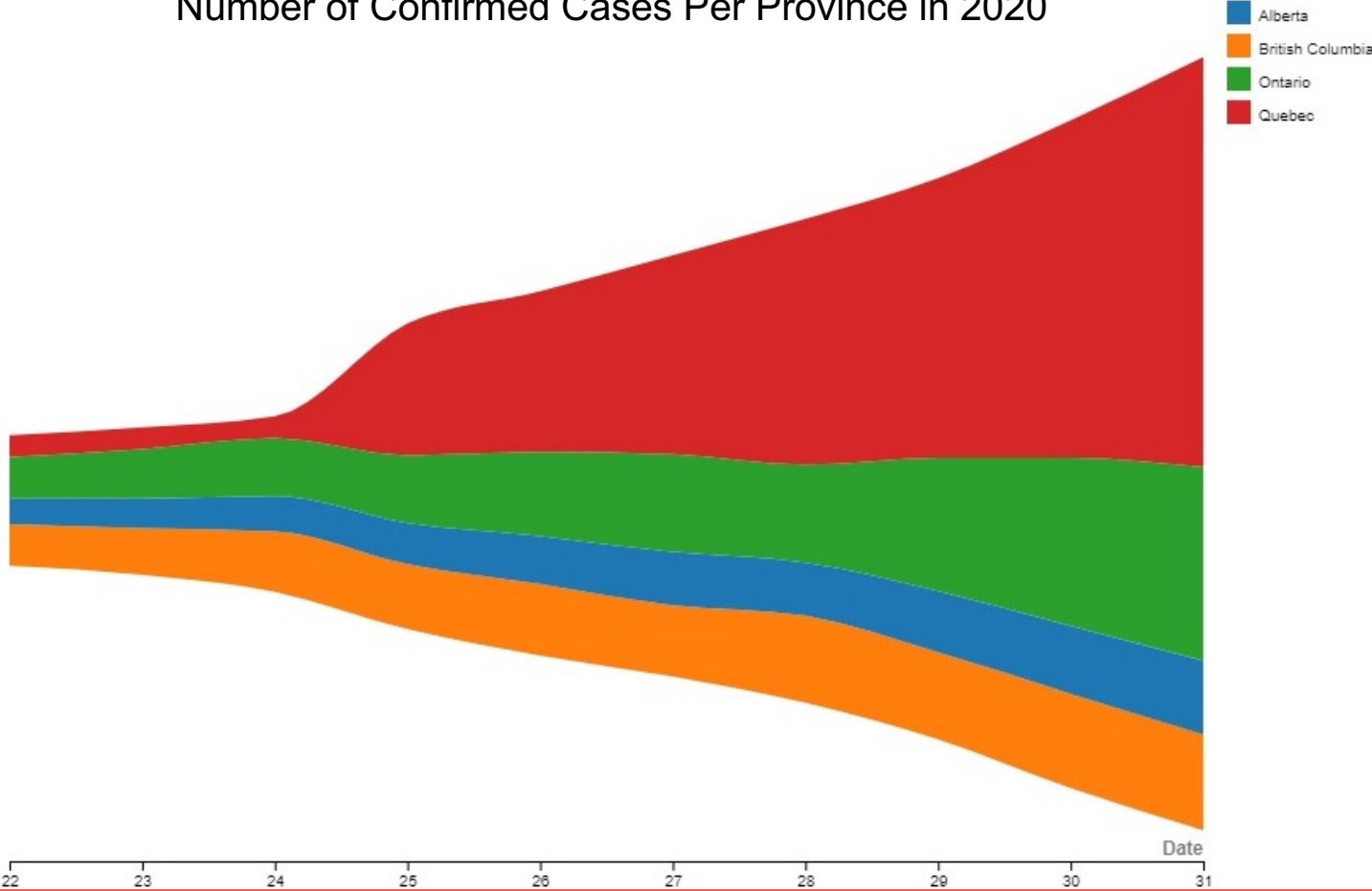 - Perceptual factors (e.g., easiest to read the group that is lowest visually on the graph)

Stacked Bar Graph

Number of Confirmed Cases Per Province in 2020

Stream Graph

Number of Confirmed Cases Per Province in 2020

Alberta
British Columbia
Ontario
Quebec

Date

22 23 24 25 26 27 28 29 30 31

Anticipated Results

- We're hoping to derive basic perceptual principles
  - e.g., confirming whether it's easier to read stream graphs when the slope is steeper (might have less of an impact on a bar graph)
  - e.g., we often impose lines to show trends or patterns in data and this might be a way that we are familiar with analyzing our data – given the continuous lines of a stream graph, might suggest they'd be better at showing trends
    - While stacked bar graphs might be better at identifying data at a specific point
- Expecting a type of question by type of graph interaction
- Specific results include
  - Position on graph, number of groups, judging proportions, etc.