Vaclav Brezina & Raffaella Bottini
Lancaster University, UK

Data visualization in Corpus Linguistics: Practical session

In this session, you will be using **the Lancaster Stats Tools online** (http://corpora.lancs.ac.uk/stats).
In addition, you will use **MS Excel** or a similar spreadsheet program to open the dataset.

Data: http://corpora.lancs.ac.uk/stats/data/Dataset_Data_Visualization_CL.xlsx

Warm-up activity: Think about how you could improve the following visual displays.

	Visual display	Notes for improvement
1	**Word cloud representing the concept of Social Media.** Source: https://studentmodules.wiki.dublinschools.net/Word+Cloud	
2	**Horizontal bar chart - The most spoken languages of the world** Source: https://conceptdraw.com/a413c3/preview	

Vaclav Brezina & Raffaella Bottini
Lancaster University, UK

Objective: In these tasks, we will become familiar with the *Graph tool* included in the **Lancaster Stats Tools online** by creating visual presentations of statistical data.
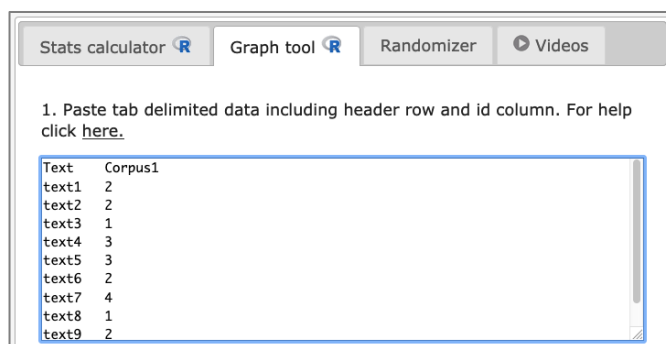
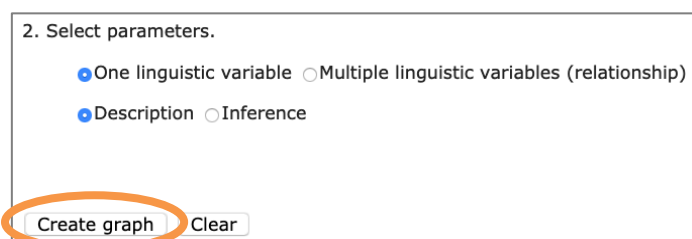### Task 1. Displaying frequency and distribution of data: histograms and boxplots

Step 1: Prepare data. Open the Excel file provided which contains the data sets. Make sure the data are in a tab-delimited format, for example, preparing the data in a spread sheet program such as Excel or Calc.

Make sure the data set includes a header row with the names of (sub)corpora and a first column with the text or speaker IDs.

Step 2: Input data. Work with "Dataset 1". Copy the data including the header row and ID column. Go to http://corpora.lancs.ac.uk/stats/toolbox.php?panel=0&tab=1 to use the *Graph tool*. Paste the data in the *Graph tool* text box.

Step 3: Select parameters and create graph.

Step 4: Interpret the graph.

   a. What type of information does the graph include?
   b. Why would you use this type of visualization? What are the advantages/disadvantages?
   c. What are the main trends and patterns in the data set?

Repeat Steps 2 to 4 with "Dataset 2".

Vaclav Brezina & Raffaella Bottini
Lancaster University, UK

Task 2. Creating different types of descriptive data visualization

Use the *Graph tool* and the data sets provided to create different types of descriptive graphs. The table below helps select the appropriate settings.

Visualization	Dataset	Parameters
Histogram	1	One linguistic variable; Description
Boxplot	2	One linguistic variable; Description
Scatter plot	4	Multiple linguistic variables; Description
Scatter plot matrix	5	Multiple linguistic variables; Description
Line chart	3	Multiple linguistic variables; Description
Geo mapping chart	6	One linguistic variable; Description
Stacked bar chart	7	One/many linguistic variable(s); Description
Sparklines	8	One/many linguistic variable(s); Description
Candlestick plot	9	One/many linguistic variable(s); Description

How can you interpret the graphs? Answer the questions in Task 1 (step 4).

Task 3. Creating inferential data visualization

Use the *Graph tool* and the data sets provided to display inferential statistics. The table below helps select the appropriate settings.

Visualization	Dataset	Parameters
Error bars	2	One linguistic variable; Inference
Scatter plot with regression line	4	Multiple linguistic variables; Inference

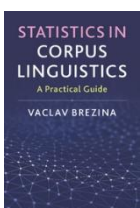How can you interpret the graphs? Answer the questions in Task 1 (step 4).

Vaclav Brezina & Raffaella Bottini
Lancaster University, UK

Optional task: Use *Graph tool* to visualize the statistical details in the table below. You can find the data sets at http://corpora.lancs.ac.uk/stats/materials.php?panel=0&tab=2

Visualization	Dataset
Histogram	'The' in BE06
	Passives in BE06 - genres
Scatter plot	'The' & 'I' in BNC64
Line chart	Modals in the Brown family - summary
Geo mapping chart	Go/travel BNC
Error bars	Passives in BE06 - genres
Scatter plot with regression line	'The' & 'I' in BNC64

How can you interpret these graphs? Answer the questions in Task 1 (step 4).

References:

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Graph tool manual: http://corpora.lancs.ac.uk/stats/docs/pdf/manuals/graph_tool_manual.pdf