



Scientific Session: BIG DATA

uRos 2021: “Network Visualization of Multi-data Sources using R”

Rui ALVES



Shirley ORTEGA-AZURDUY



Christina PIERRAKOU



GitHub



Scientific Session: BIG DATA

Visit the github page and download the R script or the html file

<https://github.com/ruialv/VizNet-uRos2021>

The context

- [ESSnet2018] ESSnet on Big Data 2018 -2020 - Eurostat grant ESTAT-PA11-2018-8 Multipurpose statistics and efficiency gains in production.

https://ec.europa.eu/eurostat/cros/content/essnet-big-data-i_en#WP7_Multiple_domains

- ESSnet Big Data II WPJ – Innovative Tourism Statistics - Task 1C

https://ec.europa.eu/eurostat/cros/content/WPJ_Innovative_tourism_statistics_en

The context



Objective

Develop a conceptual framework and setting up a prototype of Tourism Information and Monitoring System

Combining data

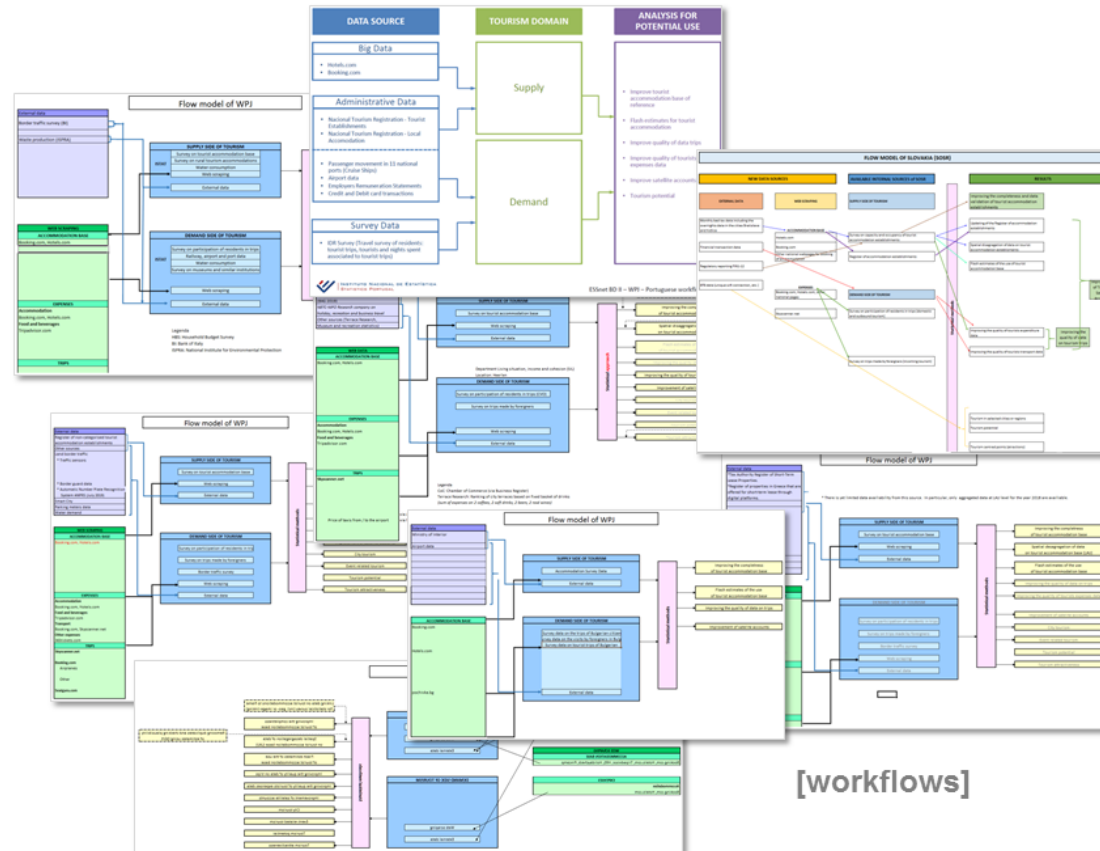
- Multi-purpose data (administrative)
- Survey data
- Web data (webscraping)

The problem

- How to get rapid overview of the data sources types
- How to trace back and forth where are sources used and to which end
 - Identify/Visualize which other countries use the same (or similar) sources
 - Understand the different purposes leading to the use of an specific source
 - Be able to directly browse into the external sources and get new insights
- Need to support production process to assess potential efficiency gains

The problem

Multiple individual workflows do not give the full picture

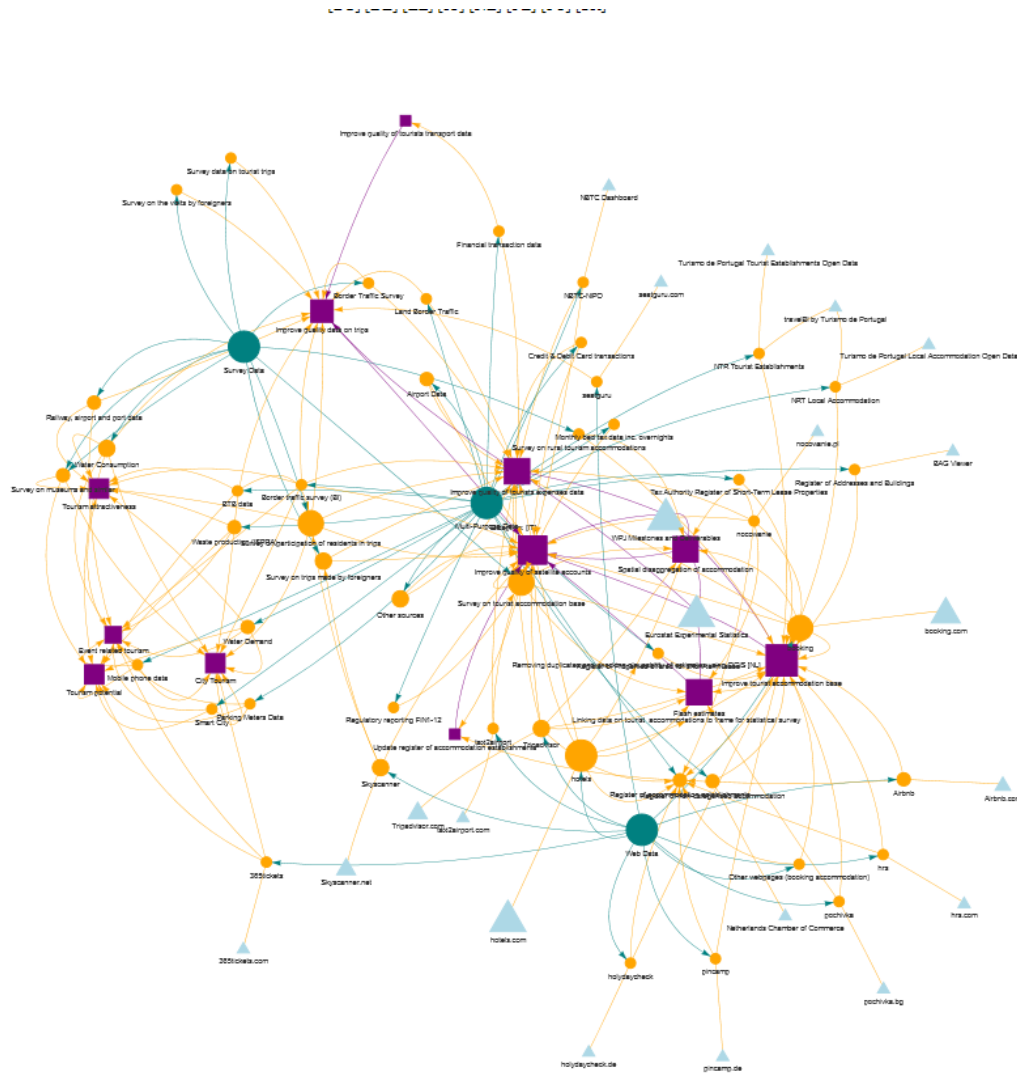


The Solution: Network Visualization

Select by node

Select by group

Edit

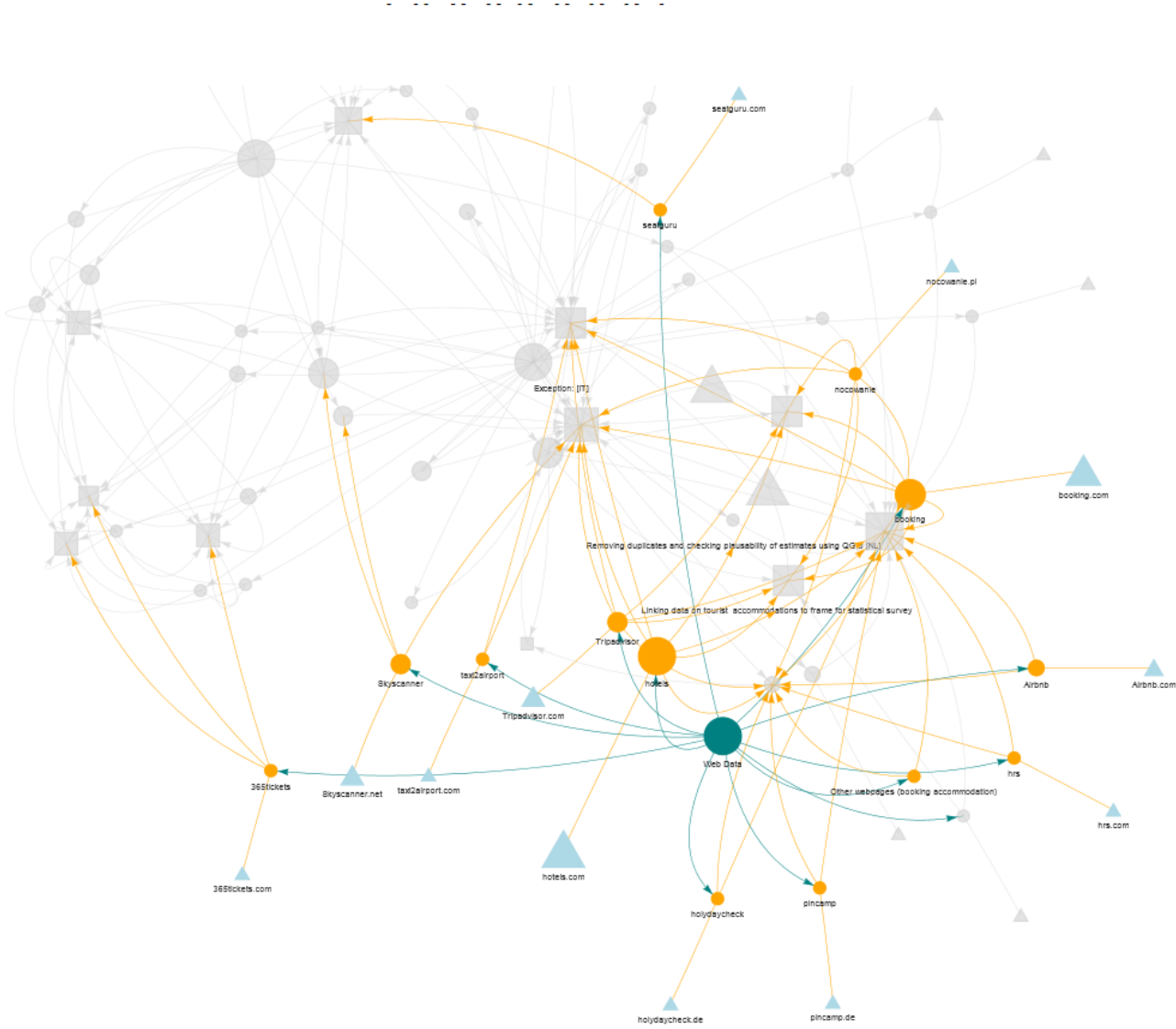


What's the importance of web scraped data?

Select by node ▾

Data Source: Web Si ▾

Edit

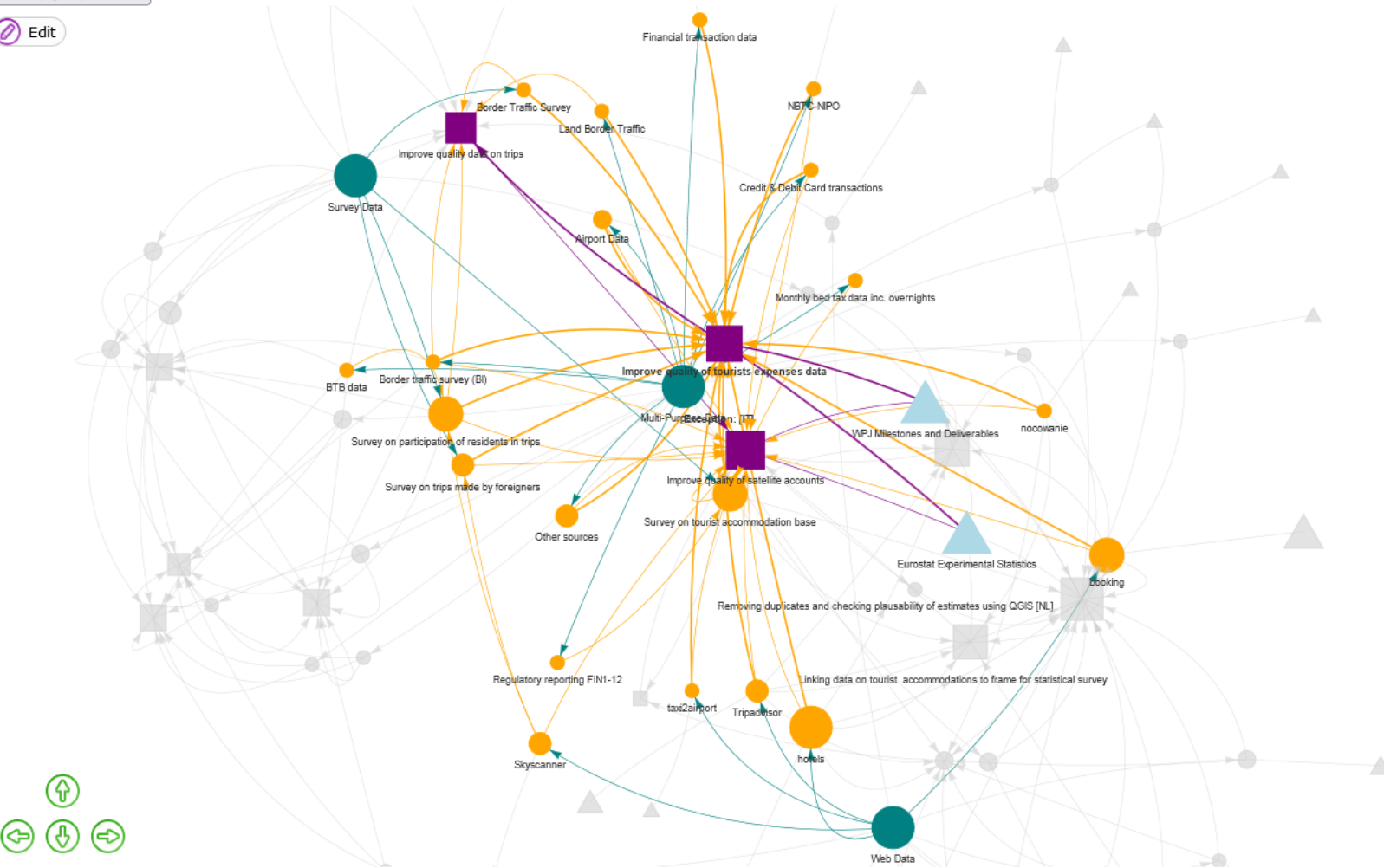


How many data sources are used to improve quality of tourist expenses data?

Improve quality of tou

Select by group

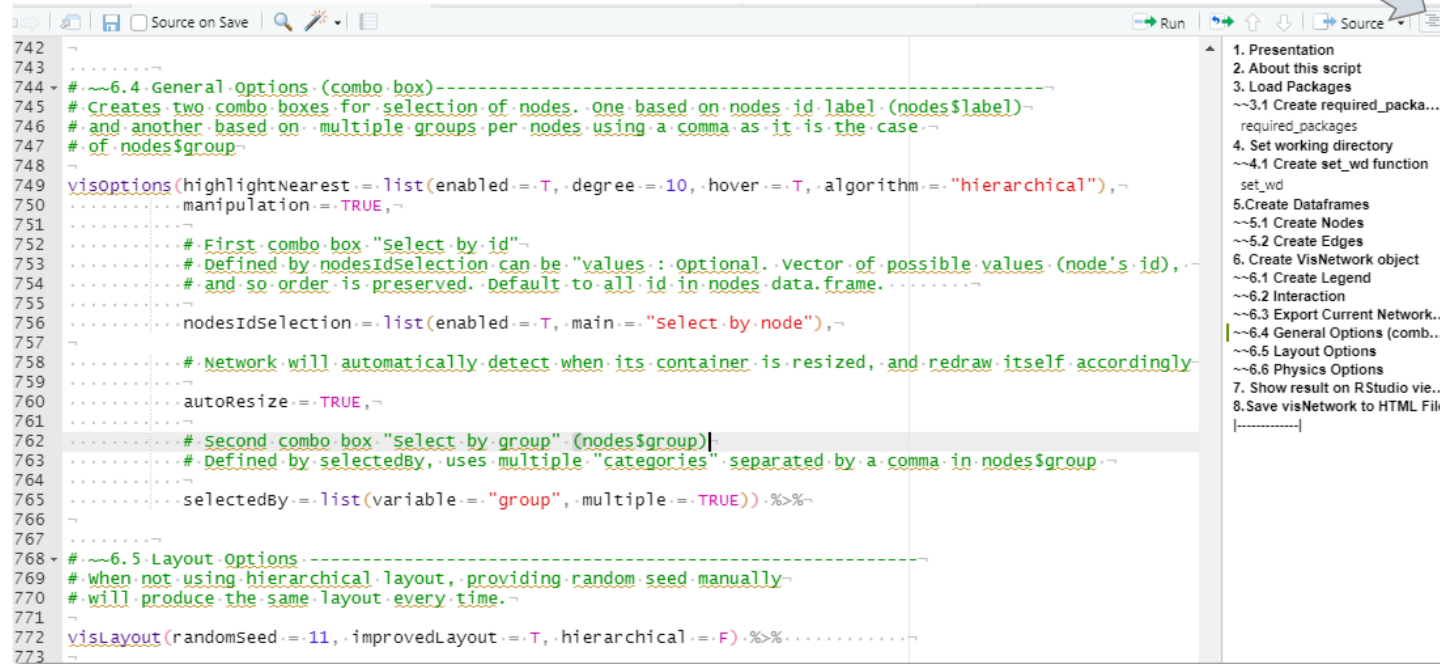
Edit



How does it work? The Script

<https://github.com/ruialv/VizNet-uRos2021>

Ctrl+Shift+O



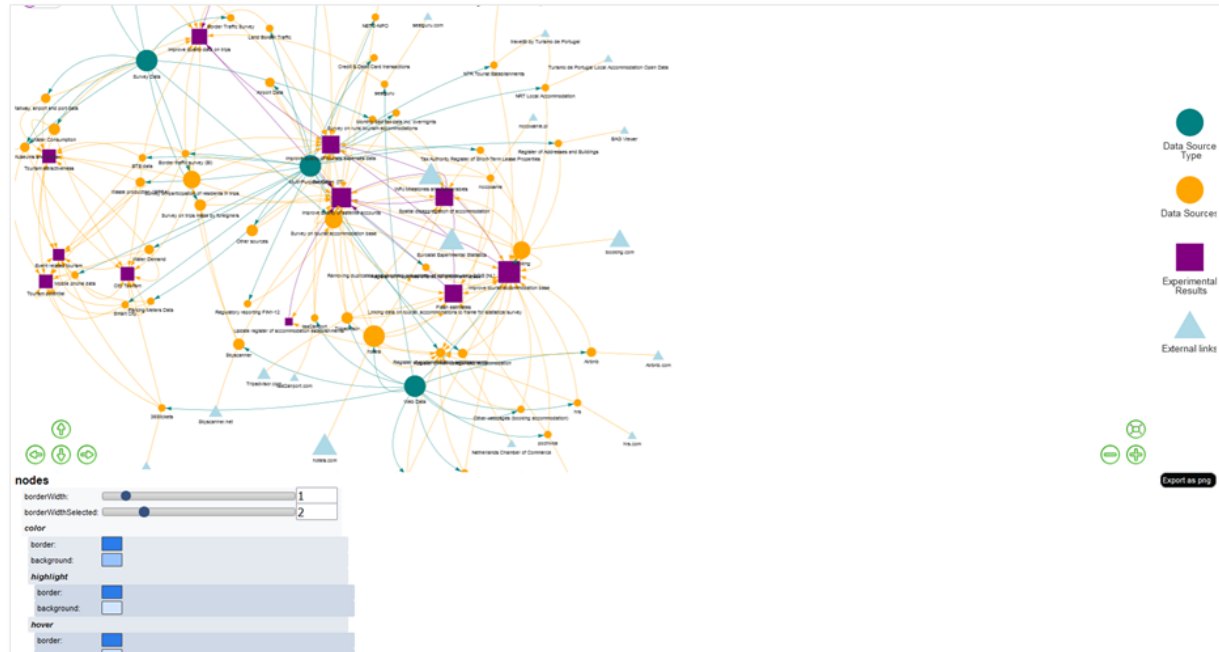
The screenshot shows the RStudio interface. The main editor window contains R code with line numbers 742 to 773. The code includes comments and function calls for configuring a network visualization. A table of contents is visible on the right side of the editor, with the entry for '6.4 General Options (comb...)' highlighted. An arrow points from the text 'Ctrl+Shift+O' to the 'Source' button in the top right corner of the RStudio window.

```
742 ~
743 ~
744 # ~6.4 General Options (combo box)-----
745 # Creates two combo boxes for selection of nodes. One based on nodes id label (nodes$id)~
746 # and another based on multiple groups per nodes using a comma as it is the case ~
747 # of nodes$group~
748 ~
749 visoptions(highlightNearest=.list(enabled=.T, degree=.10, hover=.T, algorithm=. "hierarchical"),~
750 ~~~~~manipulation=.TRUE,~
751 ~~~~~)
752 ~~~~~# First combo box "select by id"~
753 ~~~~~# Defined by nodesIdSelection can be "values : optional vector of possible values (node's id),~
754 ~~~~~# and so order is preserved. Default to all id in nodes data frame. ~~~~~
755 ~~~~~
756 ~~~~~nodesIdSelection=.list(enabled=.T, main=. "select by node"),~
757 ~~~~~)
758 ~~~~~# Network will automatically detect when its container is resized, and redraw itself accordingly~
759 ~~~~~)
760 ~~~~~autoResize=.TRUE,~
761 ~~~~~)
762 ~~~~~# Second combo box "select by group" (nodes$group)~
763 ~~~~~# Defined by selectedBy, uses multiple "categories" separated by a comma in nodes$group.~
764 ~~~~~)
765 ~~~~~selectedBy=.list(variable=. "group", multiple=.TRUE).%>%~
766 ~~~~~)
767 ~~~~~)
768 # ~6.5 Layout Options -----
769 # when not using hierarchical layout, providing random seed manually~
770 # will produce the same layout every time.~
771 ~
772 visLayout(randomSeed=.11, improvedLayout=.T, hierarchical=.F).%>%~
773 ~
```

Table of Contents:

- 1. Presentation
- 2. About this script
- 3. Load Packages
 - ~3.1 Create required_packa...
- 4. Set working directory
 - ~4.1 Create set_wd function
- 5. Create Dataframes
 - ~5.1 Create Nodes
 - ~5.2 Create Edges
- 6. Create VisNetwork object
 - ~6.1 Create Legend
 - ~6.2 Interaction
 - ~6.3 Export Current Network...
 - ~6.4 General Options (comb...)
 - ~6.5 Layout Options
 - ~6.6 Physics Options
- 7. Show result on RStudio vie...
- 8. Save visNetwork to HTML File

How does it work? visConfigure{visNetwork}



R Packages

- **{dplyr}**
 - Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7 <https://CRAN.R-project.org/package=dplyr>
- **{visNetwork}**
 - Almende B.V. and Contributors, Benoit Thieurmél and Titouan Robert (2021). visNetwork: Network Visualization using 'vis.js' Library. Rpackage version 2.1.0. <https://CRAN.R-project.org/package=visNetwork>
- **{rstudioapi}**
 - Kevin Ushey, JJ Allaire, Hadley Wickham and Gary Ritchie (2020). rstudioapi: Safely Access the RStudio API. R package version 0.13. <https://CRAN.R-project.org/package=rstudioap>
- **{shiny}**
 - Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges (2021). shiny: Web Application Framework for R. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>



Scientific Session: BIG DATA

Thank you for your attention

<https://github.com/ruialv/VizNet-uRos2021>

