How to draw the Sankey diagram and forest plot

A. The Sankey diagram

1. Step 1: getting data from the link

Download 100-top cited articles for the two journals in this study at

http://www.healthup.org.tw/html100/spine2journals.htm

2. Step 2: SNA approach to get the element counts and relations between elements
|   | A        | B        | C     | D     | E     | F       | G       | H       |
|---|----------|----------|-------|-------|-------|---------|---------|---------|
| 1 |          |          |       |       |       |         |         |         |
| 2 | C1       | 2011     | 49    | 21729796 | 1     | 2011    | U.S     |
| 3 | C1       | 2013     | 42    | 21311399 | 1     | U.S     | F5.Physiol Spine J |
| 4 | C1       | 2012     | 40    | 22045006 | 1     | F5.Physiol Spine J |
| 5 | C1       | 2014     | 34    | 22142627 | 1     | Spine J | Journal Art |
| 6 | C1       | 2015     | 24    | 23970107 | 1     | Journal Art | 21929796 |
| 7 | C1       | 2016     | 8     | 23722572 | 1     | 2012    | U.S     |
| 8 | C1       | 2017     | 3     | 21952190 | 1     | U.S     | F1.Pain & P |
| 9 | C2       | US       | 107   | 21192221 | 1     | F1.Pain & P | Spine (Phil J)
| 10| C2       | Japan    | 11    | 25839387 | 1     | Spine (Phil J) | Journal Art |
| 11| C2       | South Korea | 11  | 23537454 | 1     | Journal Art | 21311399 |
| 12| C2       | Netherlands | 9    | 24113358 | 1     | U.S     | F2.Statist & |
| 13| C2       | China     | 8     | 26208232 | 1     | F2.Statist & | Spine (Phil J) |
| 14| C2       | Canada    | 7     | 24412416 | 1     | Spine (Phil J) | Comparative |
| 15| C2       | U.K       | 7     | 23369499 | 1     | Spine (Phil J) | Journal Art |
| 16| C2       | Australia | 7     | 24239940 | 1     | 2013    | U.S     |
| 17| C2       | Germany   | 4     | 23880297 | 1     | F1.Pain & P | Spine (Phil J) |
| 18| C2       | Sweden    | 3     | 23369495 | 1     | Spine (Phil J) | Fischer Evaluation |
| 19| C3       | F2.Statist & | 74   | 24412416 | 1     | F5.Physiol Spine J |
| 20| C3       | F5.Physiol Spine J | 46  | 21192221 | 1     | Evaluation | 23970107 |

**Excel module to draw the Sankey**

To arrange data as the form below:

3. Step 3: to draw the Sankey diagram
|   | A | B | C |
|---|---|---|---|
| 1 |   |   |   |
| 2 | 0.0001 | C0 |
| 3 | C0 0.0001 | C1 |
| 4 | [49] 2011 |
| 5 | [42] 2013 |
| 6 | [40] 2012 |
| 7 | [34] 2014 |
| 8 | [24] 2015 |
| 9 | [3] 2016 |
| 10 | [3] 2017 |
| 11 | 2013 [23] F2.Statist&Date |
| 12 | 2012 [17] F2.Statist&Date |
| 13 | 2014 [9] F2.Statist&Date |
| 14 | 2015 [12] F2.Statist&Date |
| 15 | 2016 [4] F2.Statist&Date |
| 16 | 2017 [2] F2.Statist&Date |

Codes generated to fit the requirement of Sankey format at
https://sankeymatic.com/

4. Step 4: Copy and Paste them onto the Sankey maker
https://sankeymatic.com/build/
If data were pasted on the website at http://www.healthup.org.tw/kpiall/forestplot.asp, the Sankey can be drawn in this way below:
| Year | Category      | Value |
|------|--------------|-------|
| 2011 | F5 Physiology| 26.000000 |
| 2011 | F5 Physiology| 50.000000 |
| 2011 | Spine J      | 122.000000 |
| 2011 | Journal Article | 180.000000 |
| 2011 | Journal Article | 217297056 |
| 2012 | F1 Pain & Prognosis | 12.000000 |
| 2012 | F1 Pain & Prognosis | 44.000000 |
| 2012 | Spine (Phila Pa 1975) | 92.000000 |
| 2012 | Journal Article | 21311399 |
| 2012 | Journal Article | 21146297 |
| 2013 | F2 Statist & Date | 30.000000 |
| 2013 | F2 Statist & Date | 82.000000 |
| 2013 | Spine (Phila Pa 1975) | 38.000000 |
| 2013 | Comparative Study | 22045008 |
| 2013 | Journal Article | 22146297 |

For the format from different sources, see below examples...

Copy from MS Excel and Paste them onto the boxes: one for data, another for sample size, respectively. The first plot immediately appears on Google Maps.

**Forest Source:** the Sankey

| Clear |
|-------|
| Submit |
B. The Forest plot [http://www.healthup.org.tw/kpiall/forestplot.asp](http://www.healthup.org.tw/kpiall/forestplot.asp)

Data1:

- Item01 94 22 92 20
- Item02 98 21 92 22
- Item03 98 28 88 26
- Item04 94 19 82 17
- Item05 98 21 88 22
- Item06 96 21 92 22

Data2 for sample:

- 60 60
- 65 65
- 40 40
- 200 200
- 50 45
- 85 85
Data were pasted onto the website.

To confirm the parameters as shown below and click on the submit icon.

**Appendices:**

**Word cloud tutorial in Excel**
# Latent Semantics Analysis (LSA) in Excel tutorial

[https://help.xlstat.com/6698-word-cloud-tutorial-excel](https://help.xlstat.com/6698-word-cloud-tutorial-excel)

# k-means clustering in Excel tutorial

# Text Mining

- [Latent Semantics Analysis (LSA) in Excel tutorial](https://help.xlstat.com/6698-word-cloud-tutorial-excel)
- [Cleaning text data in Excel tutorial](https://help.xlstat.com/)
- [Convert text data to lower or upper case in Excel](https://help.xlstat.com/)
- [Word cloud tutorial in Excel](https://help.xlstat.com/)
- [Feature extraction tutorial in Excel](https://help.xlstat.com/tutorial-guides/text-mining)
  - [https://help.xlstat.com/tutorial-guides/text-mining](https://help.xlstat.com/tutorial-guides/text-mining)

# Feature extraction tutorial in Excel

- [https://help.xlstat.com/6751-feature-extraction-tutorial-excel](https://help.xlstat.com/6751-feature-extraction-tutorial-excel)

|   | A      | B     | C     | D    | E    | F       | G    | H   |
|---|--------|-------|-------|------|------|---------|------|-----|
| 1 | PMID   | surger | method | effect | advers | patholog | imag |    |
### Fig. Results by EFA

| view | class | F1     | F2     | F3     | F4     | F5     |
|------|-------|--------|--------|--------|--------|--------|
| 0    | 2     | -0.32101 | 0.03681 | -0.0707 | 2.33759 | -0.09017 |
| 0    | 5     | -0.39668 | -0.28528 | -0.39899 | -0.22724 | -0.33354 |
| 0    | 5     | -0.40333 | -0.19003 | -0.2521 | -0.27256 | -0.29651 |
| 0    | 2     | -0.15271 | -0.09669 | -0.57433 | -0.12948 | -0.32873 |
| 0    | 5     | -0.42663 | 0.1435 | 1.70921 | 2.36492 | -0.47225 |
| 0    | 5     | -0.60601 | -0.3051 | -0.21353 | -0.44124 | -0.49399 |
| 0    | 5     | -0.59026 | -0.26332 | -0.71941 | -0.03679 | 0.06833 |
| 0    | 5     | -0.44818 | -0.30172 | -0.31189 | -0.2248 | -0.33361 |
| 0    | 5     | -0.35832 | -0.18718 | -0.18751 | -0.0684 | -0.09843 |
| 0    | 4     | -0.5094 | -0.37439 | -1.29245 | 0.43195 | -0.36751 |
| 0    | 2     | 0.59565 | -0.00099 | -0.69276 | -0.18936 | -0.6921 |
| 0    | 5     | -0.326 | -0.21978 | -0.38035 | -0.15048 | -0.21591 |
| 0    | 2     | -0.28894 | -0.22263 | -0.77322 | 0.0557 | -0.4598 |
| 0    | 2     | -0.20421 | -0.11514 | -0.48723 | -0.12704 | -0.32861 |
| 0    | 4     | -0.5094 | -0.37439 | -1.29245 | 0.43195 | -0.36751 |

**EFA in Excel**

[https://www.real-statistics.com/multivariate-statistics/factor-analysis/]
Using the factor scores to perform the k-means clustering and select the number of clusters to five equal to the factor number.

The document related to the cluster can be obtained.
To classify articles into subject categories:

1. dataset {PMID, MeSh terms, Keywords} using SNA and determine cluster number=5 that will be clustered as below: That is, each article can be classified into each category using the SNA.

2. factor scores in two journals:

   Because each articles with each factor score have been known, the SMD can be compared using the forest plot.

   Similarly, the counts in each category for each journal can be obtained to perform the ChiSQ test for examining the count distributions equally across categories.