Data Visualisation in Spreadsheets

Julia Lesky, Majda Osmic, Stefan Simon

21 May 2014

Abstract

Spreadsheet applications are calculation and data processing applications, probably the most popular being Microsoft Excel. Considering the fact that the Microsoft Office Suite is widely used in business and home offices and that there are several other similar alternatives, it makes sense to explore the data visualisation potentials of these tools. In this survey, built-in features as well as some extension tools for Microsoft Excel and one free alternative, LibreOffice Calc, are examined. The built-in features of Microsoft Excel and LibreOffice Calc are investigated and advantages and disadvantages of both applications enumerated. Furthermore, this survey covers an overview of two plugins for Microsoft Excel, NodeXL and chart-me, one a plugin for visualisation of network data and one a tool for the creation of expressive business charts. In the last point, a short summary of a few apps for Microsoft Excel 2013, available in the Microsoft Office Store, is given.

Contents

Chapter 1

Introduction

"A picture says a thousand words." These days, data comes in a variety of forms, but cannot be interpreted easily most times. At this point, data visualisation becomes more and more important, as the human mind captures and interprets data more easily through vision than any other sense. Although spreadsheet applications provide numerous ways to process data and adapt it to the user's needs, it is often easier to just look at a diagram to see trends or get a general overview of a particular dataset. In this survey, a closer look at all possibilities to data visualisation in spreadsheet applications is taken. Microsoft Excel, a commercial spreadsheet application, provides a wide range of built-in features, plugins and add-ons, not only for data processing, but for data visualisation as well. For a direct comparison LibreOffice Calc, part of a free open source Office suite, is considered and the same examinations are done.

In Chapter 2, the built-in data visualisation features of Microsoft Excel are presented. A short introduction on data pre-processing with Pivot Tables is done, followed by some visualisation possibilities, including simple standard charts, Pivot Charts, which are standard charts linked to the data of the chosen Pivot Table, and Sparklines, which are small charts the size of one cell. Furthermore, export possibilities of visualisations are examined.

In Chapter 3, the same aspects are considered in LibreOffice Calc, a free alternative to Microsoft Office. The first thing presented is again the creation of a Pivot Table to prepare data for the following visualisations. However, for the reason that LibreOffice Calc is not as powerful as Microsoft Excel, only an overview of the standard charts is presented. The last points covered are image export of the visualisations, performance and a list of some drawbacks in comparison to Microsoft Excel.

Chapter 4 covers two plugins for Microsoft Excel. The first, NodeXL, is a plugin to download and visualise network with a broad variety of customisation possibilities to interpret network data visualisations, which makes it different from all other features, plugins and add-ons in this survey. The second, Chart-me, is a commercial tool to create templates for meaningful and consistent business dashboards. In Chapter 5, some of the apps which can be purchased in the Microsoft Office Store, are introduced.

In Chapter 6, all previous points are summarised, the presented applications compared and advantages and drawbacks listed for a final conclusion.

Disclaimer: The report was created using LaTeX based on the skeleton provided by Keith Andrews. All graphics have been created by the authors.

Chapter 2

Microsoft Excel built-in features

Microsoft Excel is the most widespread spreadsheet program. It costs about €60 in the standalone version but is also available in a package with applications like Word and PowerPoint in Office 365 from €4,10 per month [Microsoft Corporation, 2014b].

2.1 Pivot tables

An important part of creating good charts is a good data set. This could be easily created with Pivot Tables, which are built-in features of Microsoft Excel (Figure 2.1). They seem to be complicated in the first place, but the learning curve is steeply rising and there are many tutorials available [Microsoft Corporation, 2010][Video2Brain, 2014]. It is possible to connect two or more Pivot Tables. That is, all filters and group settings are synchronized.

2.2 Pivot Chart

Pivot Charts (Figure 2.2) are currently only available on Microsoft Excel for Windows. One advantage of Pivot Charts is that charts update automatically when filtering or grouping of the pivot table has changed. Pivot Charts have no restrictions to standard charts unless it isn't possible to make a diagram of the column "Grand Total". The only way to achieve this is to copy the values in a separate chart and create a standard chart or manually grouping the values to a single column (see Figure 2.3).

2.3 Conditional Formatting

Conditional Formatting is a possibility to format a cell depended of its value. Rules can be applied to change font style, background and borders. It is not possible to use relative references in these rules. There are some predefined rules, which can not be reproduced with self-created rules, for example icons directly in the cell, color scales or bar charts (Figure 2.5).

2.4 Sparklines

Sparklines are an enchancement of Conditional Formatting and let the user create small diagrams in a cell. They are limited to line charts, bar charts and Win/Loss bar charts (Figure 2.5).

2.5 Combined Charts

In Excel it is easy to combine several chart types into one chart. Very useful are the automatic calculated Trendlines (Figure 2.5).
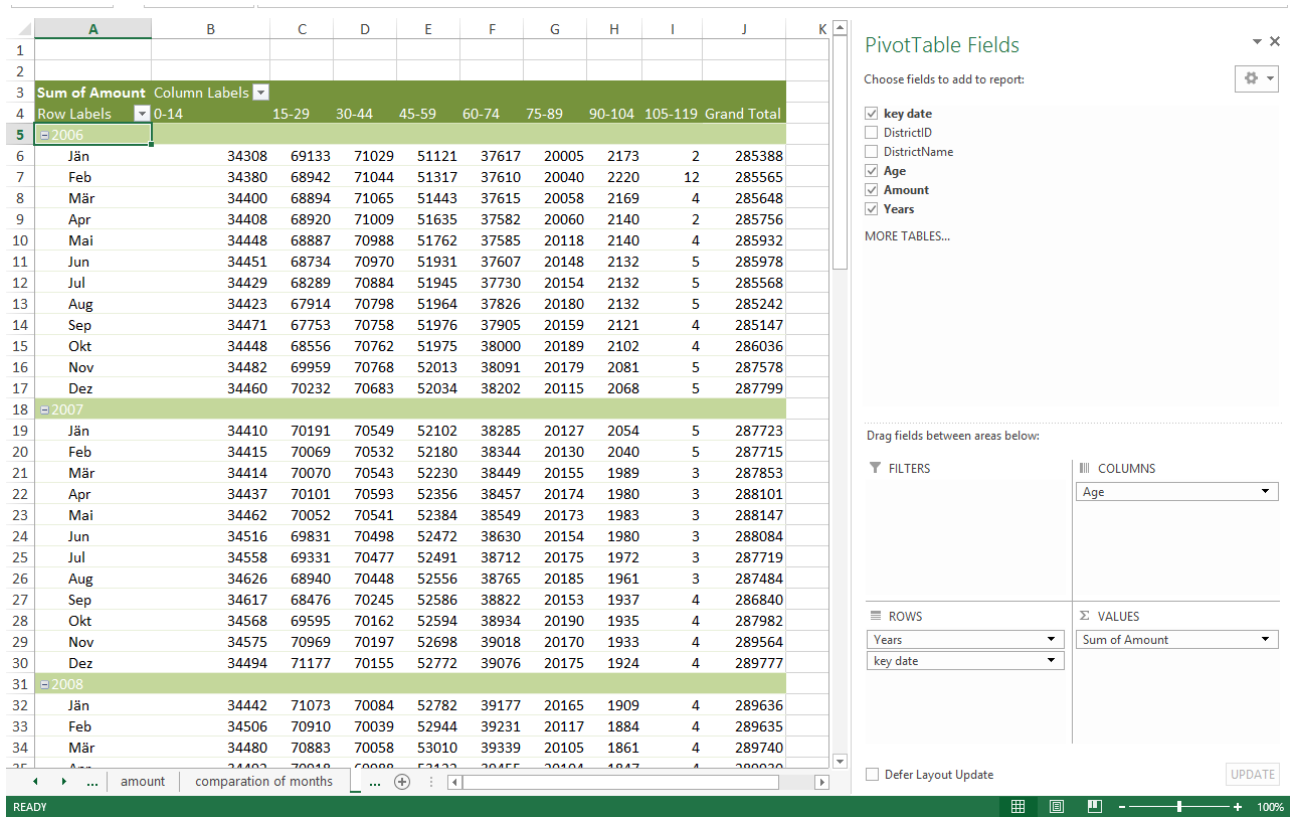
	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3	**Sum of Amount**	Column Labels									
4	Row Labels	0-14		15-29	30-44	45-59	60-74	75-89	90-104	105-119	Grand Total
5	⊟ 2006										
6	Jän		34308	69133	71029	51121	37617	20005	2173	2	285388
7	Feb		34380	68942	71044	51317	37610	20040	2220	12	285565
8	Mär		34400	68894	71065	51443	37615	20058	2169	4	285648
9	Apr		34408	68920	71009	51635	37582	20060	2140	2	285756
10	Mai		34448	68887	70988	51762	37585	20118	2140	4	285932
11	Jun		34451	68734	70970	51931	37607	20148	2132	5	285978
12	Jul		34429	68289	70884	51945	37730	20154	2132	5	285568
13	Aug		34423	67914	70798	51964	37826	20180	2132	5	285242
14	Sep		34471	67753	70758	51976	37905	20159	2121	4	285147
15	Okt		34448	68556	70762	51975	38000	20189	2102	4	286036
16	Nov		34482	69959	70768	52013	38091	20179	2081	5	287578
17	Dez		34460	70232	70683	52034	38202	20115	2068	5	287799
18	⊟ 2007										
19	Jän		34410	70191	70549	52102	38285	20127	2054	5	287723
20	Feb		34415	70069	70532	52180	38344	20130	2040	5	287715
21	Mär		34414	70070	70543	52230	38449	20155	1989	3	287853
22	Apr		34437	70101	70593	52356	38457	20174	1980	3	288101
23	Mai		34462	70052	70541	52384	38549	20173	1983	3	288147
24	Jun		34516	69831	70498	52472	38630	20154	1980	3	288084
25	Jul		34558	69331	70477	52491	38712	20175	1972	3	287719
26	Aug		34626	68940	70448	52556	38765	20185	1961	3	287484
27	Sep		34617	68476	70245	52586	38822	20153	1937	4	286840
28	Okt		34568	69595	70162	52594	38934	20190	1935	4	287982
29	Nov		34575	70969	70197	52698	39018	20170	1933	4	289564
30	Dez		34494	71177	70155	52772	39076	20175	1924	4	289777
31	⊟ 2008										
32	Jän		34442	71073	70084	52782	39177	20165	1909	4	289636
33	Feb		34506	70910	70039	52944	39231	20117	1884	4	289635
34	Mär		34480	70883	70058	53010	39339	20105	1861	4	289740

PivotTable Fields

Choose fields to add to report:

- ☑ **key date**
- ☐ DistrictID
- ☐ DistrictName
- ☑ **Age**
- ☑ **Amount**
- ☑ **Years**

MORE TABLES...

Drag fields between areas below:

▼ FILTERS	▥ COLUMNS
	Age ▼

≡ ROWS	Σ VALUES
Years ▼	Sum of Amount ▼
key date ▼	

☐ Defer Layout Update   UPDATE

... | amount | comparison of months | ... ⊕

READY   100%

Figure 2.1: Pivot Table in Microsoft Excel 2013. [Image created by the authors.]

Figure 2.2: Pivot Chart in Microsoft Excel 2013. [Image created by the authors.]

	A	B	C
1			
2			
3	**Sum of Amount**	Column Labels	
4	Row Labels	0-110	Grand Total
5	⊟ 2006		
6	Jän	285388	285388
7	Feb	285565	285565
8	Mär	285648	285648
9	Apr	285756	285756
10	Mai	285932	285932
11	Jun	285978	285978
12	Jul	285568	285568
13	Aug	285242	285242
14	Sep	285147	285147
15	Okt	286036	286036
16	Nov	287578	287578
17	Dez	287799	287799
18	⊟ 2007		
19	Jän	287723	287723
20	Feb	287715	287715
21	Mär	287853	287853
22	Apr	288101	288101
23	Mai	288147	288147

Figure 2.3: Pivot Chart with manually created "Grand Total" column. [Image created by the authors.]

Sum of Amount	Column Labels						
Row Labels	2006	2007	2008	2009	2010	2011	2012
Andritz	228301	227952	230588	230926	231386	234115	237255
Eggenberg	238631	242047	244107	244195	245260	246043	247533
Geidorf	329594	333007	338009	337109	335550	333584	336066
Gösting	130665	131621	132143	132931	133183	134281	137278
Gries	346556	345432	345226	343450	345272	351271	351788
Innere Stadt	51181	50773	49502	50307	49686	49706	51602
Jakomini	415527	418025	419459	421367	418193	421801	429529
Lend	352923	361317	367370	368673	371208	378130	388059
Liebenau	159784	161046	162540	164045	165773	168257	170303
Mariatrost	116711	118164	120495	122076	123138	124524	126289
Puntigam	92384	93117	92866	92832	92931	93074	94843
Ries	78805	78710	78179	77675	77936	78408	79035
St.Leonhard	204250	206110	208785	211035	210713	212880	214597
St.Peter	183964	184130	184820	184450	185968	188363	189688
Straßgang	169416	171887	172735	173737	176094	179733	182780
Waltendorf	155508	155756	155848	156977	156327	156672	157266
Wetzelsdorf	177437	177895	179711	180665	180514	181937	183708
Grand Total	3431637	3456989	3482383	3492450	3499132	3532779	3577619

Figure 2.4: Applying Conditional Formatting and Sparklines in Excel. [Image created by the authors.]
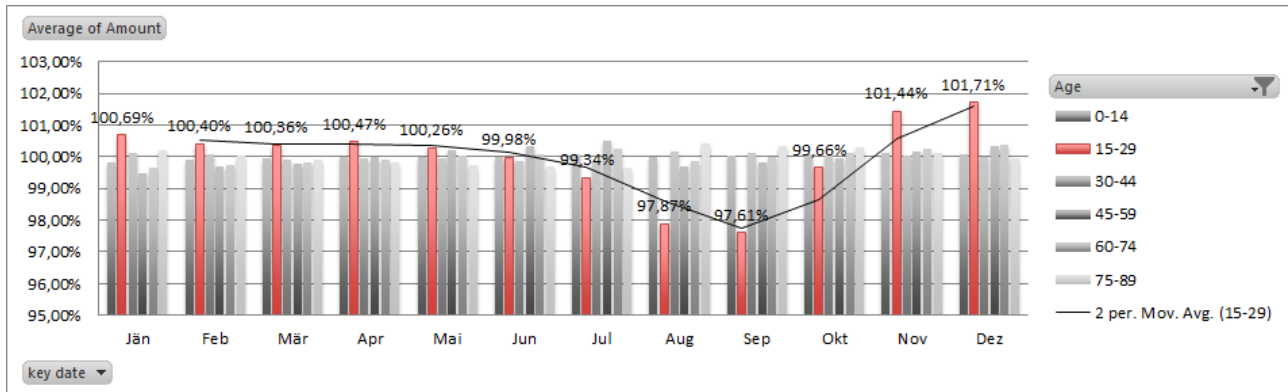
Figure 2.5: Combined Charts in Excel. [Image created by the authors.]

2.6   Export

It is possible to resize the picture arbitrary and copy it as an image. Export as vector graphic is not possible.

Chapter 3

Libre Office Calc

LibreOffice is advertised as "a powerful office suite which embeds several applications that make it the most powerful Free & Open Source Office suite on the market" [The Document Foundation, 2014a]. For that reason the LibreOffice suite is often mentioned as a free alternative to some applications of Microsoft Office.

In this chapter a closer look at LibreOffice Calc, the spreadsheet application of LibreOffice, will be taken, and, in more detail, especially built-in features and external plugins for spreadsheet data visualisation. After some research, other than for Microsoft Excel, no concrete and easy-to-use external plugins for LibreOffice Calc were found. There exists a plugin called GeOOo [The Document Foundation, 2014b], which maps numerical data to geographical areas, but it cannot be used with LibreOffice Calc alone – LibreOffice Draw is needed to process spreadsheet data and draw the visualisation. In addition to GeOOo being only able to visualise geographical data, several tests on a Windows 8 laptop with current technology to map sample spreadsheet data to a geographical map led to a crash of LibreOffice Draw. Moreover, GeOOo is not compatible with the current version number 4.2.4 of LibreOffice (by date May 18, 2014).

Therefore the focus was switched to examine only the built-in features to visualise spreadsheet data with LibreOffice Calc and compare them to results obtained with Microsoft Excel. At a first glance, it seems that LibreOffice Calc provides the same chart types as Microsoft Excel, which are the following, each with several sub-types to choose from (see Figure 3.1):

- Column Charts,

- Bar Charts,

- Pie Charts,

- Area Charts,

- Line Charts,

- XY Scatter Charts,

- Bubble Charts,

- Net Charts,

- Stock Charts and

- Column and Line Charts.

Taking a closer look, these charts are the only possibilities to visualise data in LibreOffice Calc, other than Microsoft Excel, which offers a broader variety of built-in features, like Pivot Charts, which are directly linked to Pivot Tables, or Sparklines (see Chapter 2).
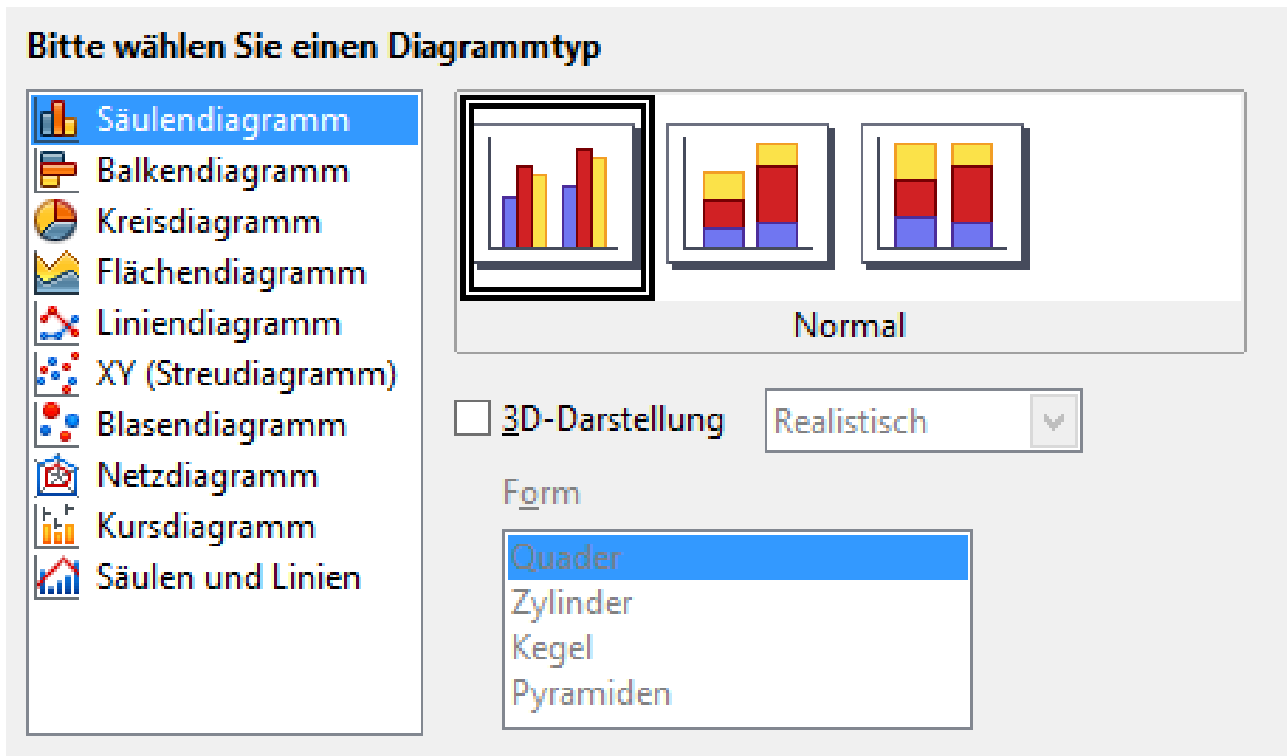
Figure 3.1: Chart types provided by LibreOffice Calc. [Image created by the authors.]

3.1 Pivot Tables

For the test charts for comparison with Microsoft Excel, as far as possible, the same dataset with the population of Graz over the last 8 years was taken from Open Government Data Graz, 2014. The dataset consists of about 150 000 records of the number of persons of every age, sorted by month of the last 8 years and in second order by district. First of all, like in Microsoft Excel, the data needs to be pre-processed to get sensible charts, which was done with the usage of Pivot tables (called Data Pilot in previous versions of LibreOfficeCalc), where data can be filtered and sorted the way it is needed. With the help of these tables, the population of Graz can be sorted by date and age. Furthermore, the data is sorted by age groups of 15 years to get clearer charts (see Figure 3.2).

3.2 Charts

Unlike in Microsoft Excel, there is no possibility to link charts to any other data range than a chosen (marked) cell range. That means it is not possible to choose the created Pivot table and link the chart to the whole range, meaning the chart will change too if changes in the Pivot table are made (like applying a new grouping or limit the data). However, for the visualisation example, the chosen cells are the dates and age groups and the chart type is a line chart. With the chart wizard, a few adjustments regarding the data range or which labels should be displayed can be applied, but the functions are mostly limited to that. However, the different areas of the chart (data rows, axes, legend, colours...) can be customized with a right click on the corresponding area. As the chart is corresponding to the chosen cell range, it would be also possible to display the total amount of people living in Graz in the same chart. Nonetheless it is displayed in another chart (Figure 3.4) so that Figure 3.3 is not distorted because of the larger data range on the y-axis. There were attempts to visualise the data in other formats, but that would have required extensive data preparation, as the chart wizard has only quite simple features to choose from for mapping data into one of the given chart types.

For an overview of more possibilities to draw and format the different chart types, Chapter 3 of the Libre-

	A	B	C	D	E	F	G	H	I	J
1	Filtern									
2										
3	Summe - Anzahl	Alter ▼								
4	Stichtag ▼	0-14	15-29	30-44	45-59	60-74	75-89	90-104	105-119	Gesamt Ergebnis
5	2006-01-01	34308	69133	71029	51121	37617	20005	2173	2	285388
6	2006-02-01	34380	68942	71044	51317	37610	20040	2220	12	285565
7	2006-03-01	34400	68894	71065	51443	37615	20058	2169	4	285648
8	2006-04-01	34408	68920	71009	51635	37582	20060	2140	2	285756
9	2006-05-01	34448	68887	70988	51762	37585	20118	2140	4	285932
10	2006-06-01	34451	68734	70970	51931	37607	20148	2132	5	285978
11	2006-07-01	34429	68289	70884	51945	37730	20154	2132	5	285568
12	2006-08-01	34423	67914	70798	51964	37826	20180	2132	5	285242
13	2006-09-01	34471	67753	70758	51976	37905	20159	2121	4	285147
14	2006-10-01	34448	68556	70762	51975	38000	20189	2102	4	286036
15	2006-11-01	34482	69959	70768	52013	38091	20179	2081	5	287578
16	2006-12-01	34460	70232	70683	52034	38202	20115	2068	5	287799
17	2007-01-01	34410	70191	70549	52102	38285	20127	2054	5	287723
18	2007-02-01	34415	70069	70532	52180	38344	20130	2040	5	287715
19	2007-03-01	34414	70070	70543	52230	38449	20155	1989	3	287853
20	2007-04-01	34437	70101	70593	52356	38457	20174	1980	3	288101
21	2007-05-01	34462	70052	70541	52384	38549	20173	1983	3	288147
22	2007-06-01	34516	69831	70498	52472	38630	20154	1980	3	288084
23	2007-07-01	34558	69331	70477	52491	38712	20175	1972	3	287719
24	2007-08-01	34626	68940	70448	52556	38765	20185	1961	3	287484
25	2007-09-01	34617	68476	70245	52586	38822	20153	1937	4	286840
26	2007-10-01	34568	69595	70162	52594	38934	20190	1935	4	287982

Figure 3.2: LibreOffice Calc Pivot Tables with age groups of 15 years. [Image created by the authors.]

Office Documentation Calc Guide can be considered [The Document Foundation, 2014c].

3.3 Export

LibreOffice provides a simple-to-use feature to export created charts. With a right click on the chart and choosing "Export as graphic" the saving dialogue is opened. Common graphic types like PNG or JPEG are supported and it is also possible to export the chart as a vector graphic (SVG).

3.4 Performance

With an older version of LibreOffice on a Windows 8 laptop with current technology, several problems were encountered. Even with this comparably small dataset LibreOffice Calc was pushed to its limits. Processing data, like marking a certain data range and then just clicking on the chart wizard or even saving files led to corresponding times of several minutes or to a crash of the whole application in several cases. The same visualisation steps were applied later in LibreOffice Calc on Linux as well as on Windows 8 with the current version of LibreOffice, 4.2.4. Although LibreOffice Calc is still remarkably slower with a dataset of this size than Microsoft Excel, especially operations like search/replace or saving, no severe performance issues were encountered again.

3.5 Remarks

At this point some drawbacks of the free spreadsheet application LibreOffice in comparison to the commercial Microsoft Excel can be listed:

- Datasets need extensive pre-processing to meet the simple requirements of the given chart types which are provided by LibreOffice Calc.

- LibreOffice Calc does not handle exhaustive datasets very well. Even a relatively small set with 150 000 records slows down the whole application remarkably and leads even to a crash of the application in
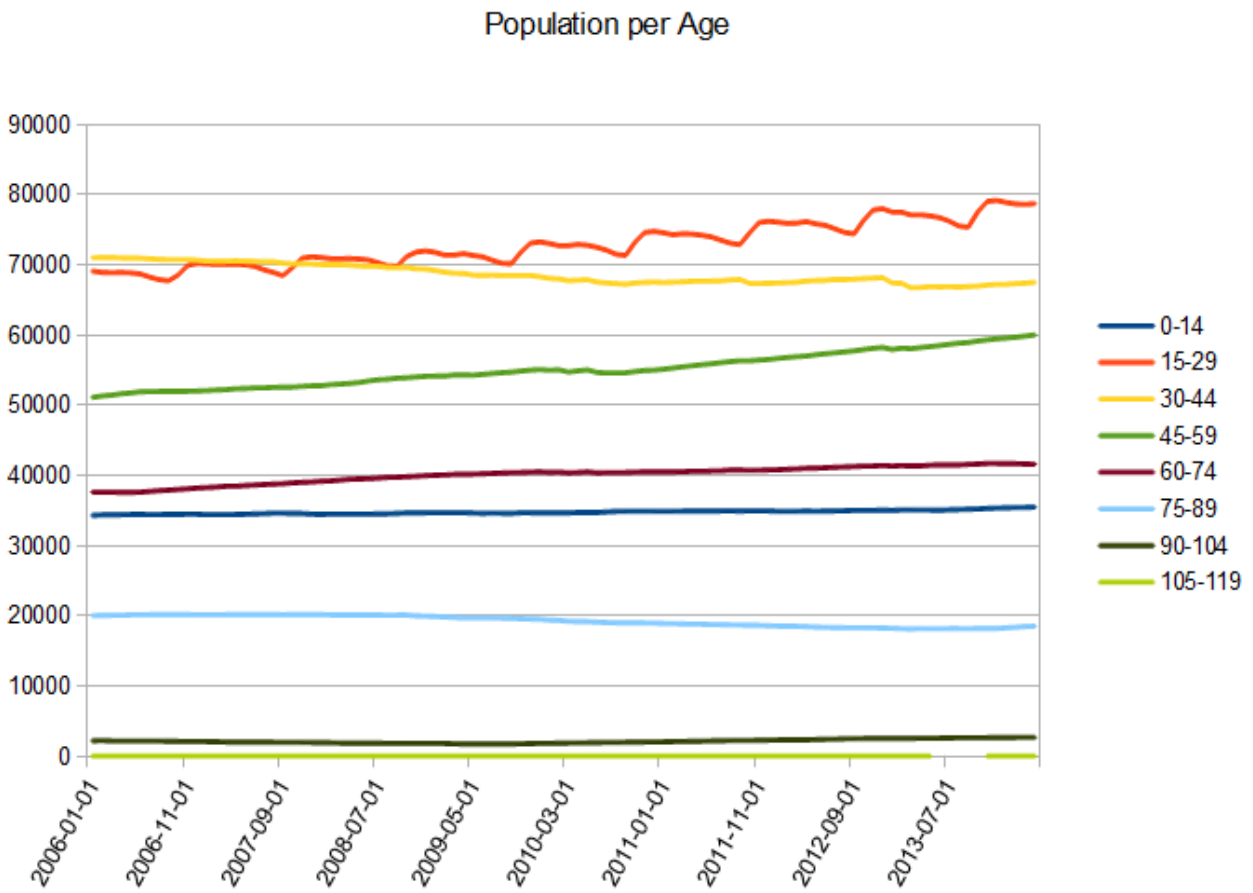
Figure 3.3: Line Chart with one line for every age group, except the total amount. [Image created by the authors.]
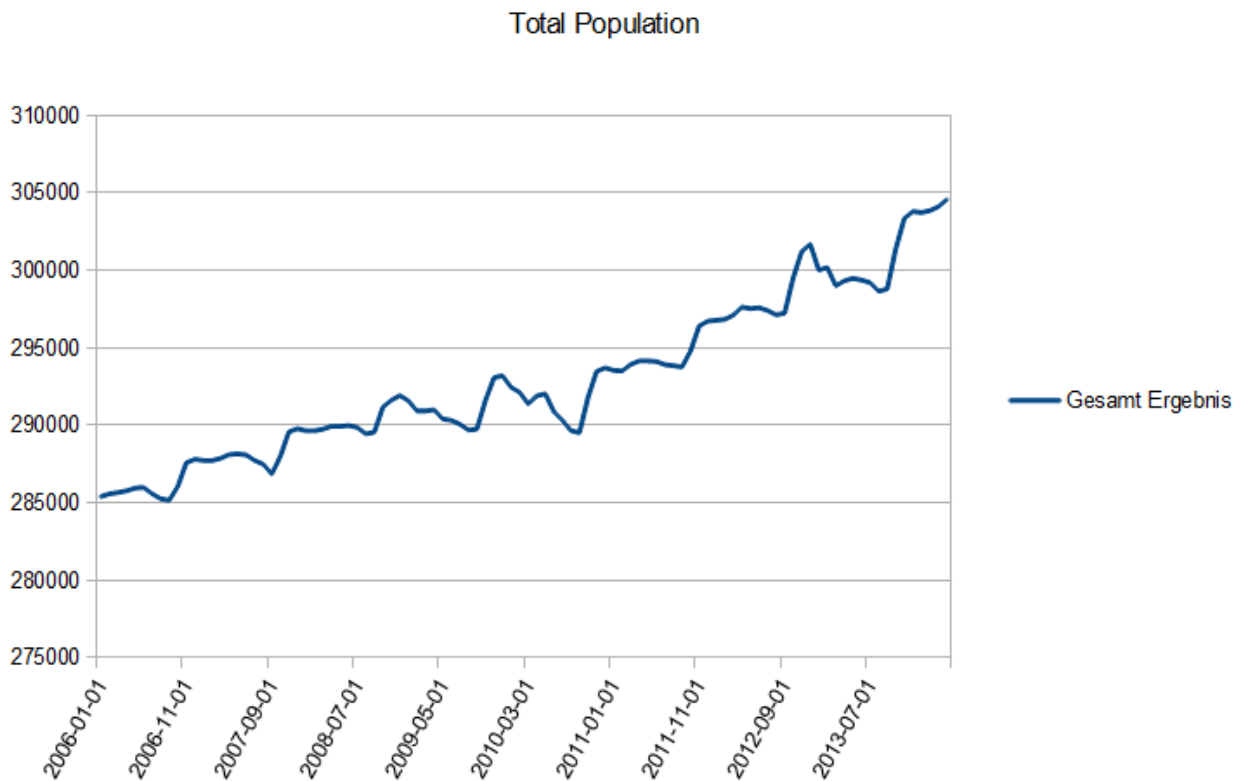
Figure 3.4: Line chart of the total population since 2006.l [Image created by the authors.]

some cases, especially when files should be saved. However, an update to the current version results in a great improvement of the overall performance.

- As the charts are only simple charts, they are only dependent on the chosen cell range. Unlike Microsoft Excel, there is no way to make the charts dependent on a whole Pivot table and therefore to change dynamically when elements are filtered in the Pivot table. The chosen cell range stays the same.

- There are no extensions or plugins which can be used for more powerful data visualisations.

All things considered and even if LibreOffice Calc has its limitations, the application is a useful and free alternative to commercial Spreadsheet applications. It provides a basic set of visualisation possibilities to draw simple charts for either simple or pre-processed data of small range. Therefore, LibreOffice Calc is a sufficient solution for small and exemplary data visualisations.

Chapter 4

Microsoft Excel Plugins

4.1 NodeXL

NodeXL is a Microsoft Excel 2007, 2010 and 2013 add-in for interactive visualisation and analysis of network graphs [CodePlex, 2014a]. The creation of the NodeXL add-in was supported by the Social Media Research Foundation [Social Media Reserach Foundation, 2014a] whose main goal is to develop tools optimised for collection, visualisation and analysis of social media network data. NodeXL is free and open-source.

4.1.1 Template

NodeXL is an add-in for Microsoft Excel 2007 or higher. When installed, NodeXL adds a new section in the Microsoft Excel Ribbon. A spreadsheet template is also provided with several sheets that define the list of vertices, edge list (directed connections between the vertices), groups and overall network metrics. Columns in each sheet contain network data and parameters for visual representation of vertices and edges, such as label, colour, width, style, opacity, and many other visual and other attributes.

NodeXL provides certain interactive features, such as zooming, scaling, selecting or deselecting one or more vertices along with the corresponding edges, as well as the navigation through large graphs using the hand tool. The data can be dynamically filtered in or out of the graph.

4.1.2 Data Creation and Import

The network data can be either imported or created by the user by inserting the necessary data in the template. When it comes to importing, NodeXL provides a variety of possibilities. Apart from standard network data representation formats, such as GraphML or Pajek, the data can be imported into NodeXL from any currently opened Microsoft Excel workbook in which case the user has to manually select vertex and edge columns. NodeXL can also import network data based on the emails a user has exchanged, provided that an email client such as Microsoft Outlook or Thunderbird is used. Although NodeXL can visualise any sort of network graphs, it has been primarily created for visualisation and exploration of social networks. For this reason, NodeXL provides download and import of network data from several popular social networks such as Flickr, Twitter or YouTube. Import from Facebook is possible using a Social Network Importer library[CodePlex, 2014b] for NodeXL. In order to use this functionality, an active account on the relevant social network is required. The import from Twitter can be fairly slow because Twitter has limited the number of API requests to 15 within a 15 minutes window [Twitter, 2014], so the download is often paused and continued after 15 minutes if it requires more than one API request to complete. The download of the network data from all other previously mentioned social networks runs without major difficulties .
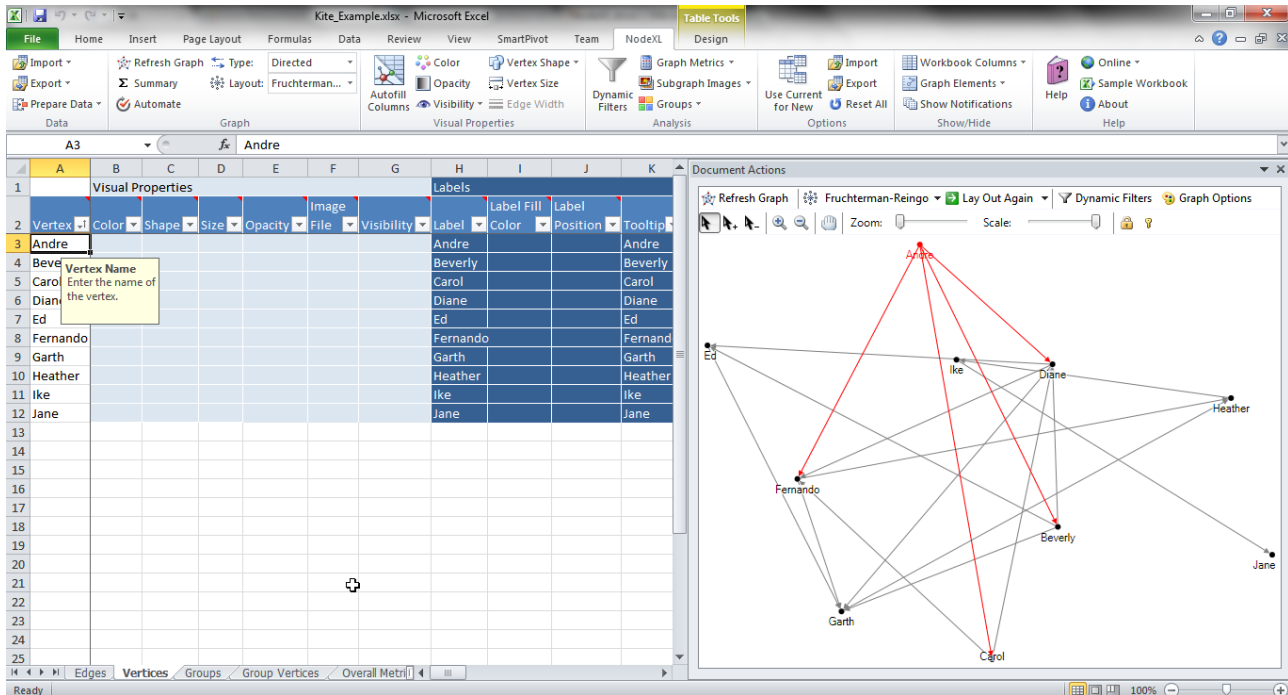
Figure 4.1: NodeXL Template displaying sample list of vertices with selected vertex and corresponding edges being highlighted. [Created with the Kite Dataset [Social Media Reserach Foundation, 2014c]]

4.1.3   Network Graph Layout

NodeXL provides different algorithms for automatic placement of vertices. Fruchterman-Reingold is one of the three force-direced graph layouts available in NodeXL. Force-directed algorithms[Wikipedia, 2014a] are iterative. The vertices are laid out based on attractive and repelling forces between them. The vertices are moved around in each iteration and the forces between them change. The goal is to achieve the minimal energy of the system. NodeXL allows users to set the strength of the repulsive force between the vertices as well as the number of iterations.

Another force directed layout algorithm available in NodeXL is Harel-Koren Fast Multiscale. This algorithm creates aesthetically pleasing layouts in which all edges have approximately the same length and the number of crossings is minimized. The algorithm is also designed to be very fast, and can create a layout for 1000 vertex graphs in about 2 seconds [Harel and Koren, 2001].

In the Circle layout all vertices are placed in the form of a circle. When using this layout it is particularly useful to apply the edge bundling algorithm which is provided in NodeXL as an additional option. This algorithm groups similar edges into bundles in order to avoid visual cluttering.

The vertices can also be laid out it in form of spiral, vertical or horizontal sine wave and grid. A layout in polar coordinates can also be created. Furthermore, the graph can be drawn using Sugiyama layered graph layout. Layered graph algorithms tend to draw network graphs in layers with consistent edge orientation and minimal crossings [Wikipedia, 2014b]. The layout which places the vertices randomly on the screen is also provided by NodeXL.

4.1.4   Group Metrics

Based on the available network graph data, NodeXL can calculate and insert a variety of graph metrics into the workbook [Hansen, Shneiderman, and Smith, 2010]:

- Vertex degree (undirected graphs) – the number of other vertices a particular vertex is connected to.
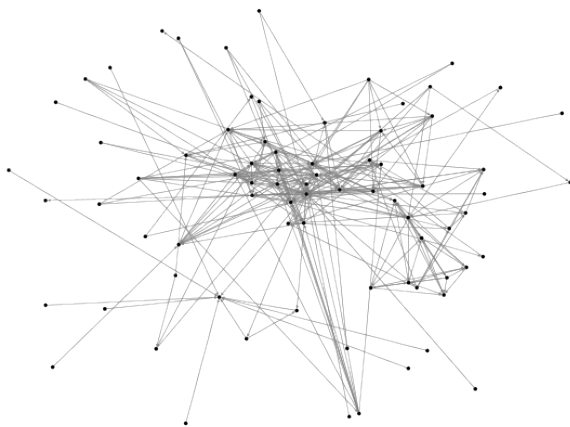
(a) Fruchterman-Reingold graph layout
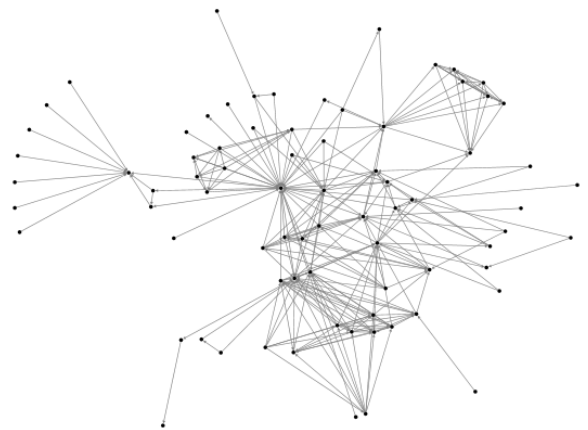
(b) Harel-Koren graph layout.

Figure 4.2: The same dataset displayed in two different force directed graph layouts available in NodeXL
[Created with the "Serious Eats" Dataset [Social Media Reserach Foundation, 2014c]]
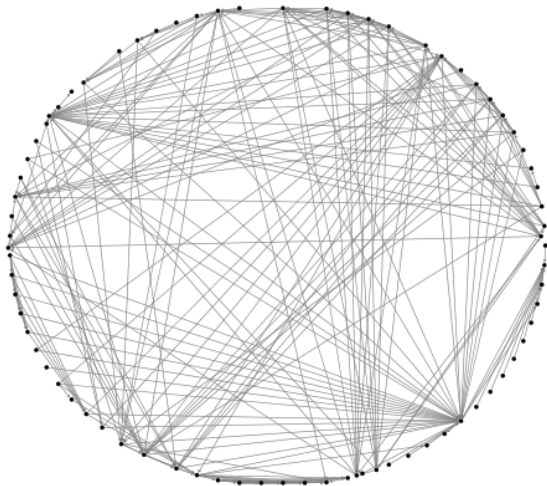
(a) Circle layout with straight edges
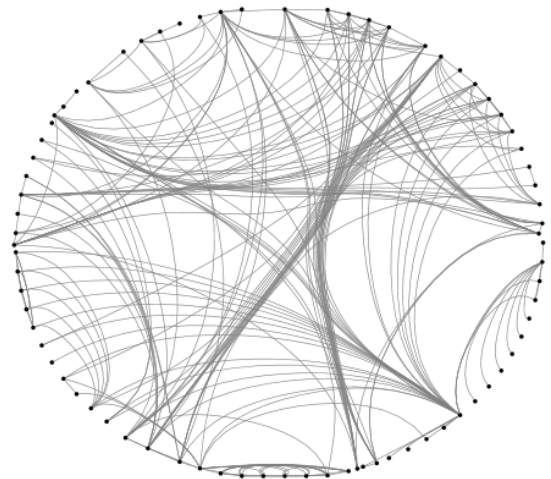
(b) Circle layout with tightly bundled edges.

Figure 4.3: If the edges in the circle layout are bundled the graph becomes much less cluttered and more
readable [Created with the "Serious Eats" Dataset [Social Media Reserach Foundation, 2014c]]
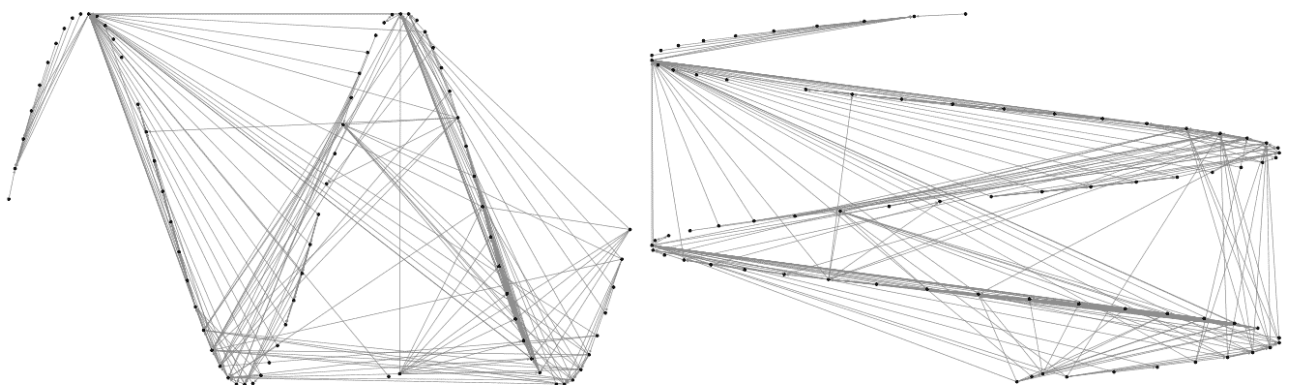
Figure 4.4: Horizontal (left) and vertical (right) sine wave graph layouts
[Created with the "Serious Eats" Dataset [Social Media Reserach Foundation, 2014c]]
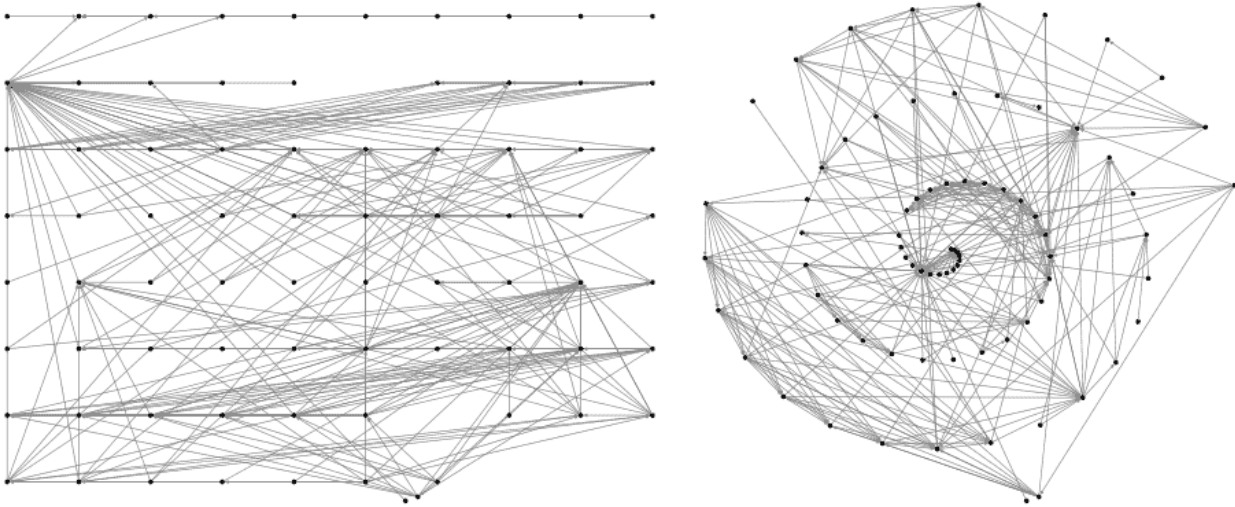
Figure 4.5: Grid (left) and spiral (right) graph layouts [Created with the "Serious Eats" Dataset [Social Media Reserach Foundation, 2014c]]

- Vertex in- and out-degree (directed graphs) – the number of other vertices that point to or from a particular vertex.

- Vertex betweenness centrality – sum of all shortest paths between nodes that pass through a particular node.

- Vertex closeness centrality – the average distance of each node to every other node in the network.

- Vertex eigenvector centrality – sum of centrality values of every node a particular node is connected to .

- Vertex page ranking (directed graphs) – the number of edges pointing towards a specific node divided by the number of all outgoing edges of each node pointing to that node.

- Vertex clustering coefficient – total number of edges connecting vertex's neighbours divided by the number of all possible edges that could connect vertex's neighbours.

- Edge reciprocation (directed graphs),

- Group metrics,

- Twitter search network top items.

4.1.5   Visual Parameters of Edges and Vertices

An important aspect of visual analysis of graphs is the differentiation of edges and vertices. In NodeXL, vertices can be visually presented in different colour, size, shape, opacity and visibility values. Vertex labels can be placed alongside of the vertex, as a pop-up that appears when the mouse is hovered over the vertex, the entire vertex can be have the shape of the label surrounded by a rectangle, or the label can be completely hidden. Edges can be represented using different colour, width, arrow size (in case of directed graphs), curvature and opacity values. Labels can be placed on edges as well. Different visual attributes can also be defined for currently selected edges and vertices. The values of visual attributes can be entered both manually into the template or using the „Auto-fill Column" functionality in which the user can define visual parameters for every calculated graph metric.
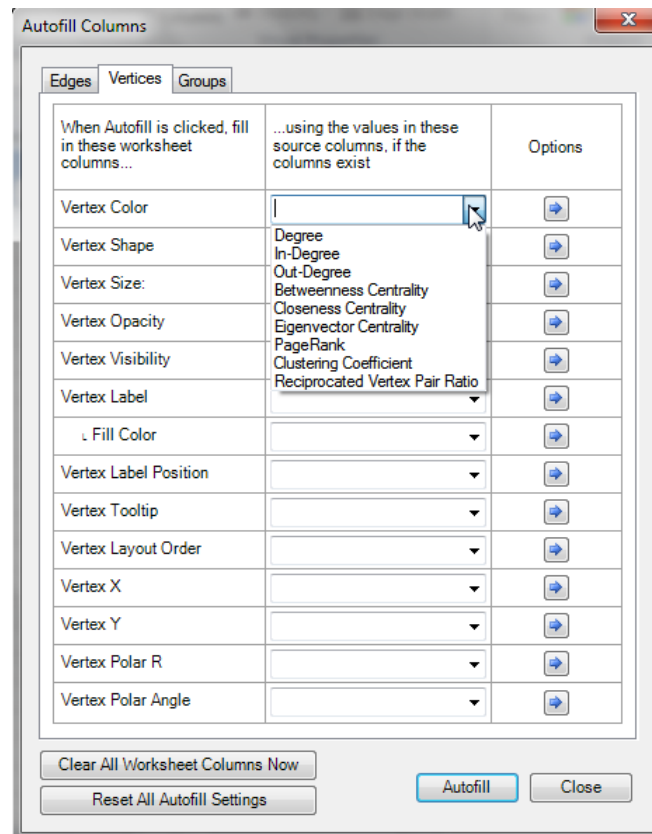
Figure 4.6: Using Auto-fill Columns option visual parameters can be set based on different graph metrics
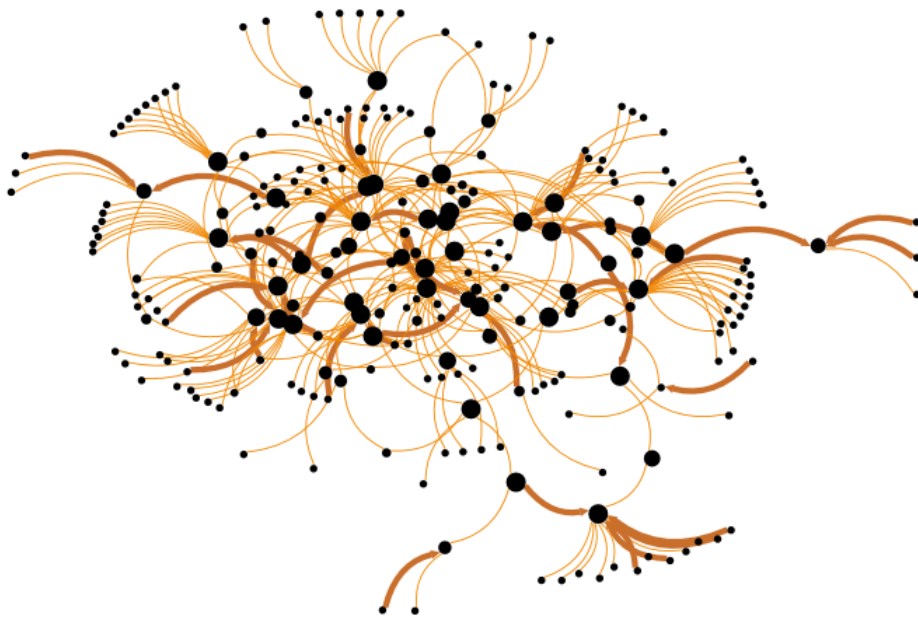
Figure 4.7: Using the Auto-fill Graph options the width of the graph edges is set to depict the edge weight and the size of the vertex depicts the betweenness centrality [Created with the "Serious Eats" Dataset [Social Media Reserach Foundation, 2014c]]
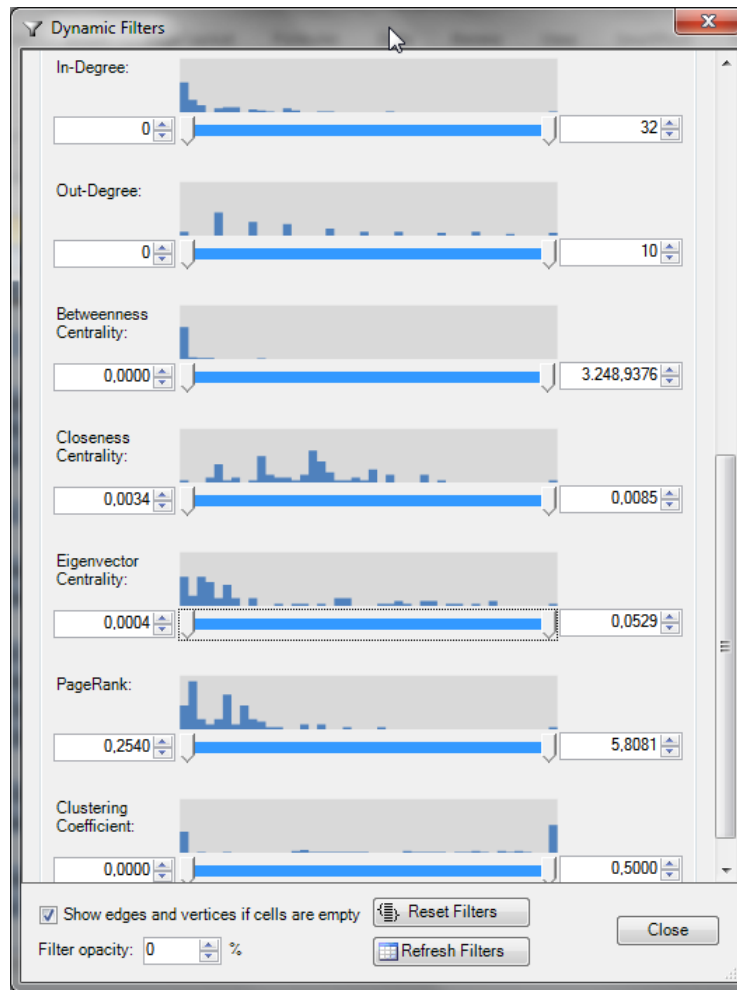
Figure 4.8: Edges and vertices can be filtered in or out of the graph based on different graph metrics using the displayed sliders

4.1.6   Data Filtering

NodeXL provides the option for auto counting and merging of duplicate edges. After the duplicate edges have been merged, the edge weight of every merged edge increases depending on how many times the edge was duplicated. Dynamic filters can be applied in order to filter out certain parts of the graph. For example, one might want to view only the vertices with certain in-degree values. Edges and vertices can be filtered based on all available graph metrics. The parts of the graph that are filtered out are not completely removed from the graph but made invisible. Because of this, the graph can be redrawn very fast even after the parts that have been filtered out are put back into the graph.

4.1.7   Grouping

In network graph analysis often groups and clusters of vertices are observed. In NodeXL, vertices can be grouped by vertex attributes such as in- and out-degrees, betweenness, closeness and eigenvector centrality, page rank and other. Custom vertex attributes can be defined and used to group the vertices. NodeXL can also identify motifs and connected components of the network graph and group the vertices accordingly. In network graphs, clusters are made of a set of densely connected vertices with only a limited number of connections to other clusters [Hansen, Shneiderman, and Smith, 2010]. NodeXL provides three algorithms for vertex clustering: Caluset-Newman-Moore, Wakita-Tsurumi, Girvan-Newman.
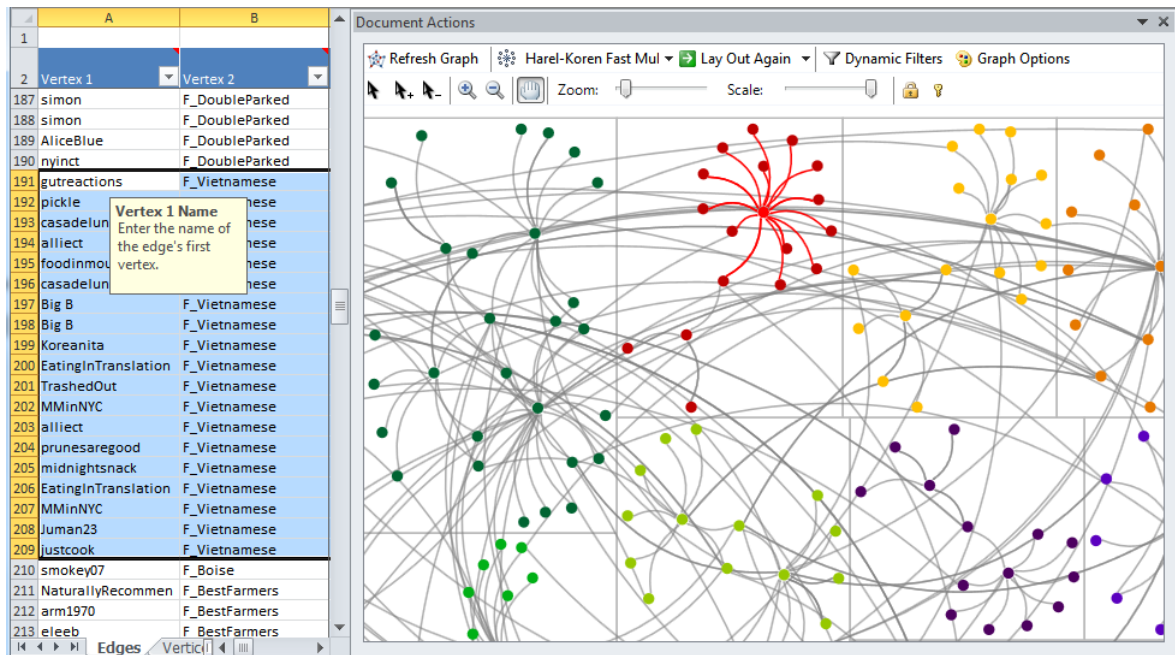
Figure 4.9: "Serious Eats" dataset [Social Media Reserach Foundation, 2014c] binds the users of the www.seriousouseats.com to the topics they have contributed to. The graph in this figure has been clustered using Wakita-Tsurumi algorithm. The highlighted edges depict all the users that have contributed to a forum topic "Vietnamese". These users all belong to the same group.

4.1.8   Subgraphs

NodeXL includes a sub-graphing feature using which sub-graphs for each vertex can be calculated. The sub-graphs depict how a particular vertex is connected other vertices and it can be calculated for different adjacency levels from 1.0 to 4.5. The subgraphs can be displayed separately for each vertex in the "Vertices" sheet of the NodeXL template or they can be exported as images into a folder on the local hard disk.

4.1.9   Graph Export and Upload to NodeXL Gallery

The graphs created with NodeXL can be exported to BMP, GIF, JPEG, PNG, TIFF and XPS formats. Additionally, the created graphs can be uploaded directly to the NodeXL Graph Gallery [Social Media Reserach Foundation, 2014b] which contains images of uploaded graphs along with all available information about the graphs (used layout, clustering algorithm, graph metrics...) which is added automatically by NodeXL and uploaded along with the image. For some graphs basic interactivity is provided, but this is momentarily just experimental functionality.

4.1.10   Performance

NodeXL has a relatively good performance for smaller datasets. For larger datasets that contain 5000 edges or more the dynamic filters tend to either freeze or completely crash the program, but other functionalities work without any problems.

4.2   Chart-me

Chart-me is a Microsoft Excel 2007, 2010 and 2013 add-in developed by Hi-Chart GmbH.[HI-CHART, 2014]. Chart-me provides extensive functionality for creation of charts optimized for use in business reports and pre-
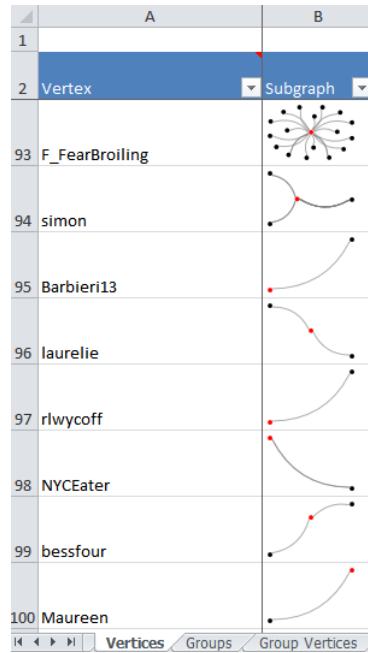
Figure 4.10: Calculated subgraphs for some vertices in the "Serious Eats" dataset. The adjacency level is
1.5 [Created with the "Serious Eats" Dataset [Social Media Reserach Foundation, 2014c]]
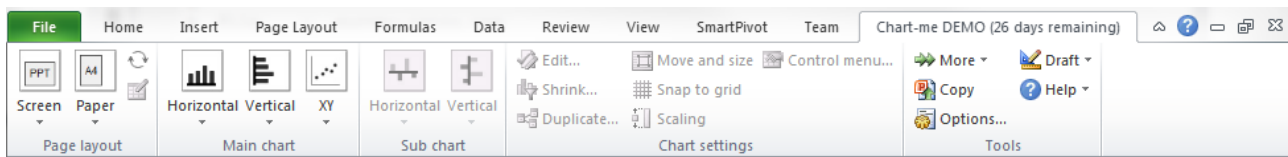

Figure 4.11: Chart-me section in Microsoft Excel Ribbon provides access to all Chart-me functionalities.

sentations. It is neither free nor open-source: a single Chart-me license costs €595.00, whereas a package with 5 user licenses costs €2,380.00. However, a free 30 days trial is available.

4.2.1   Template

When installed, Chart-me adds a new section into the Microsoft Excel Ribbon. Depending on whether the chart is being created for display on paper or screen, different chart templates can be selected. Both templates for display on screen and paper can contain one chart as well es several charts contained in 2x2, 3x3, 4x4 grids or in a tree view. For display on paper, templates optimized for A3 and A4 paper sizes are provided. When selecting a template for display on screen, the following resolutions can be selected: 4:3, 16:9, 16:10 (optimized for Microsoft PowerPoint) or custom.

After the template is created, the sheet displays an area in which the chart (or charts) will be drawn, along with a table in which the users can modify parameters such as the chart width and height, and chart labels.

4.2.2   Charts, Sub-charts and Data Insertion

Chart-me provides a variety of different types of bar charts, as well as some line and XY charts. Bar and line charts can be displayed horizontally and vertically.

Each chart available for selection is usually used for particular types of data and number of data dimensions. Based on this, whenever a chart is created, a table with sample data is also inserted into the template. The amount of data that will be visualised in the chart is selected in a dialogue upon the chart insertion and the number of columns in the generated table corresponds to the selected amount. For example, if a simple bar chart is created with 12 columns in the data insertion table (which is the default value), the columns will
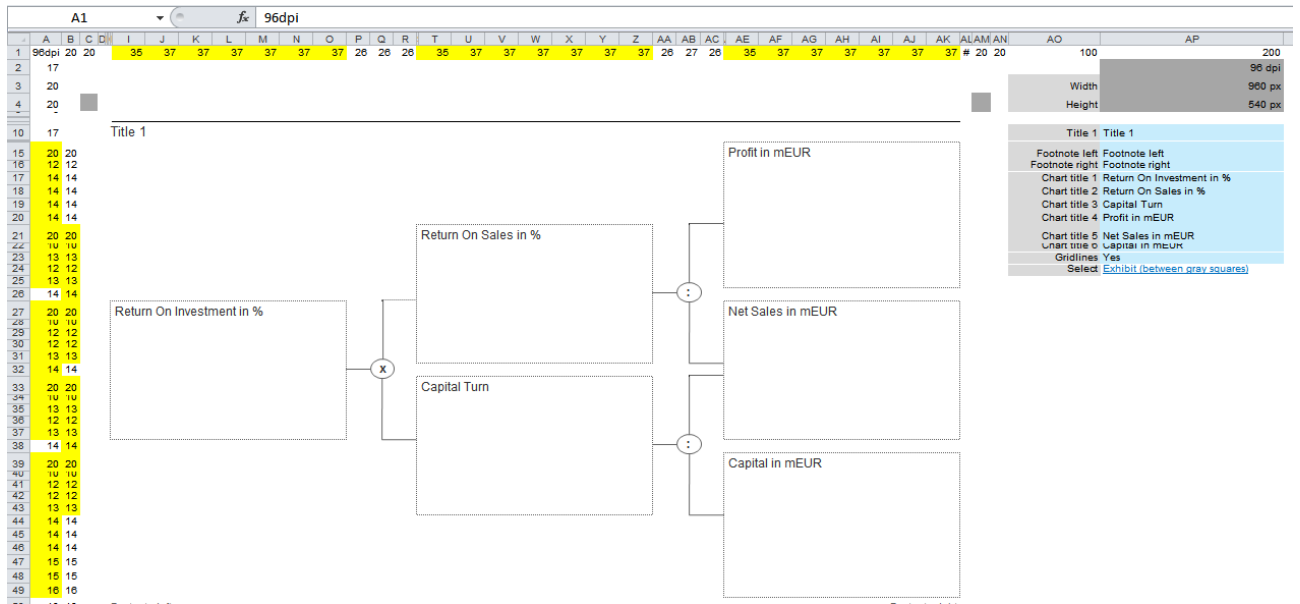
Figure 4.12: Template created by Chart-me for chart display on screen. The charts are drawn in the area on the left-hand
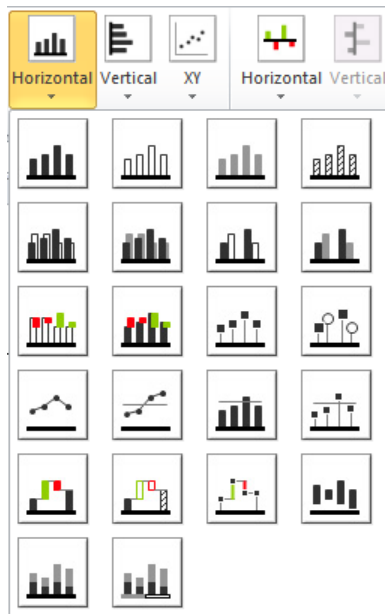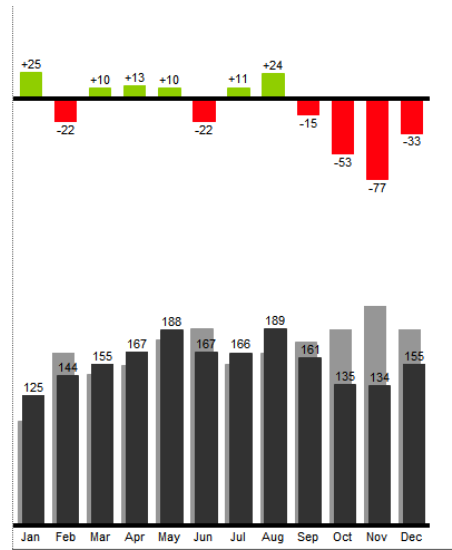

Figure 4.13: Different selection of bar charts created

(a) The generated bar chart

(b) The data insertion table

Figure 4.14: The data is read from the table and visualised the chart and sub-chart. The first four contain the data visualized in the chart, the last two rows contain values based on which the subchart displays an increase or decrease in value.

contain sample data related to 12 months in a year. The users need to modify the generated content of the table in order to display their own data. The number of input columns can be changed at any time. To each inserted chart a sub-chart can be added, which is particularly useful if an additional data dimension needs to be displayed, such as an increase or decrease in value. The built-in chart types can also be used in combination with Chart-me specific charts.

Visual parameters of the charts such as colours, line thickness, category widths and others can be modified in a special sheet provided by the Chart-me. The customized styles are applied to all charts that are created after the style modification.

4.2.3 Chart Export

Chart-me provides no chart export features. The charts created with layouts optimized for the display on the screen can be copied to a clipboard and pasted to Microsoft Power Point, whereas the charts with paper layouts can be printed directly from Microsoft Excel.

4.2.4 Performance

The insertion of charts into the template can take up few minutes and some chart modification operations can also be relatively slow even though the datasets that are being visualised with Chart-me are normally not very large. However, a progress dialogue is always shown when such operations are being executed and the program does not tend to crash. Apart from that, the charts of a particular type, font and chart size are stored in a template directory and do not need to be created every time. This means that if the same kinds of charts are reused, the

process of creation takes long only the first time.

Chapter 5

Microsoft Excel 2013 Apps

With Excel 2013, Microsoft introduced an easy way to add apps to their Office products. The Office Store can be accessed directly from Excel, and apps can be added by clicking a button. One of the main differences between Microsoft Excel plugins and apps is that upon installation of plugins, such as NodeXL or Chart-me, usually add a new section in the Ribbon, whereas the apps provide access to all their functionality through a special window displayed directly in the spreadsheet. Another difference is that the plugins are mostly available for Microsoft Excel 2007, 2010 and 2013, whereas the apps are available only for Microsoft Excel 2013.

There is a wide selection of free and and paid apps, some of which are used to visualise data. These visualisation apps mostly work only with special types of datasets that have to be inserted in a predefined data format. In the following sections, some of these apps are presented.

5.1 Office Store Visualisation Apps Overview

5.1.1 SmartCharts

SmartCharts ([devscope, 2014]) is a Microsoft Excel app that provides visualisation of date change over time. The app is available for €4.99, but a trial version that visualises only sample data can be downloaded for free. Using the free samples, a short overview of the SmartCharts app can be given.

SmartCharts can display data changes over time using bubble, doughnut, bar and tree-map charts. After the data is entered and the chart type is selected, data trends over time can be animated by clicking on the "Play" button. When the data is visualised using the bubble chart, the biggest bubble is always displayed in the centre of the chart meaning that during the animation, the bubbles do not only shrink and expand but also move around the screen. Because all these transformations happen at the same time, the chart becomes very hard to read and understand. The same problem is observed with other chart types available in the SmartCharts app.

5.1.2 Histogram

Histogram ([Microsoft Research, 2014a]) is a free Microsoft Excel 2013 app. It has been developed by Microsoft Research and it can be used to display the distribution of a data column in a histogram. Apart from colour adjustment, no other customization options are provided. Sample data is available in order to show how the app can be used. All the functionalities that are provided in this app are also available as Microsoft Excel built-in features.

5.1.3 Histogram 2D

Histogram 2D ([Microsoft Research, 2014b]) is another free app developed by Microsoft Research. It displays the distribution of two data columns in a two-dimensional histogram. Again, as a customization feature, colour can be adjusted, but no other features are provided. Sample data for the usage demonstration is available.
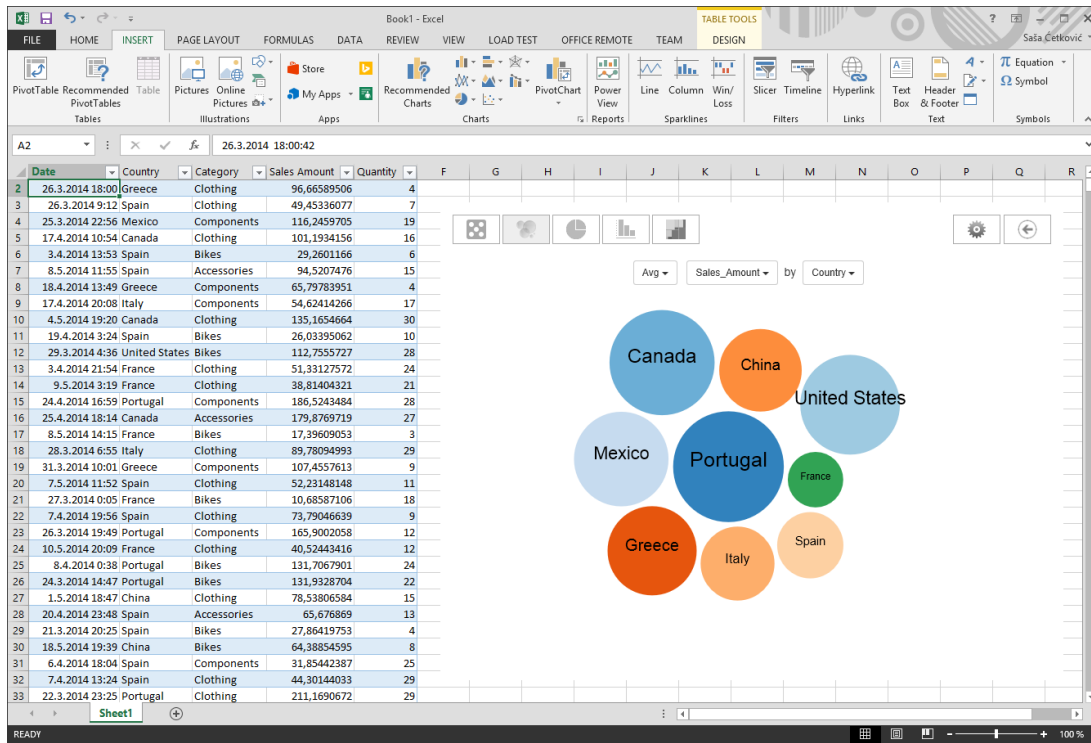
Figure 5.1: The SmartCharts sample data set and the visualisation using the bubble chart. The biggest bubble always takes the central position.

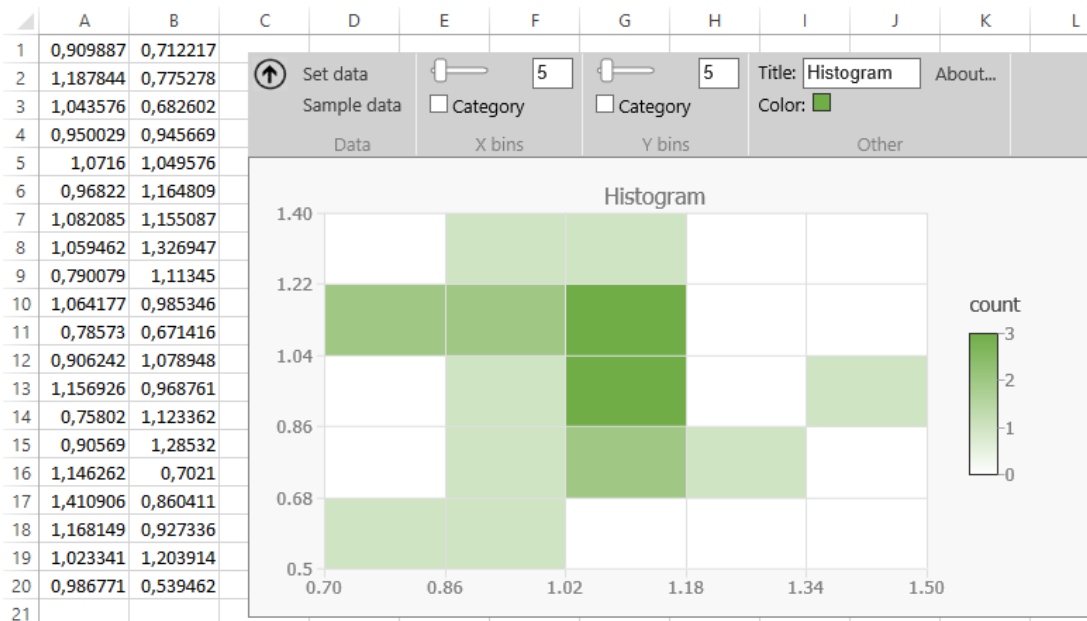**Figure 5.2:** The Histogram sample data set and the corresponding visualisation.

Figure 5.3: The Histogram 2D sample data set and the corresponding visualisation.

5.1.4 Treemap

Treemap ([Microsoft Research, 2014d]) is another free Excel app developed by Microsoft Research. In a treemap, the data is displayed in rectangles, and the size and colour of the rectangle are mapped to some data dimension. In the Treemap app, the user can map the visual parameters such as label, size, color etc. to a dataset simply by clicking on a corresponding button in the menu and selecting a cell range. The colours are automatically mapped, but the user can select what colours should be used for the mapping.

5.1.5 Streamgraph

Another free app in the set of data visualisation apps provided by Microsoft Research is the Streamgraph ([Microsoft Research, 2014c]). This app displays data changes over a period of time in a form of a stream. The horizontal axis contains the time data, the columns are mapped to the colour and actual data values are mapped to the thickness of the stream. This app automatically recognizes the label of the columns and displays them correctly in the graph.

5.1.6 Bing Maps

The free Bing Maps app ([Microsoft Corporation, 2014a]) can map multi-dimensional numerical values to circles or small pie charts into a Bing Maps widget. If the correct address format is used, a bubble can be displayed on the corresponding place in the map. The displayed bubble can contain numerical values provided by the users. This app also provides some customization options, such as colour adjustments.

5.1.7 Geographic Heat Map

With the Geographic Heatmap app ([Patel, 2014a]) it is possible to map the range of numerical values to a colour of a country in a US map. The country name needs to be spelled correctly in order to map the values. Unfortunately, there is no way to set labels on the map. This app is free.

Figure 5.4: A treemap visualisation using the Treemap Excel app. The rectangle labels are mapped to the values in the first column, the size is mapped to the values second column and the colour is mapped to the values in the third column.
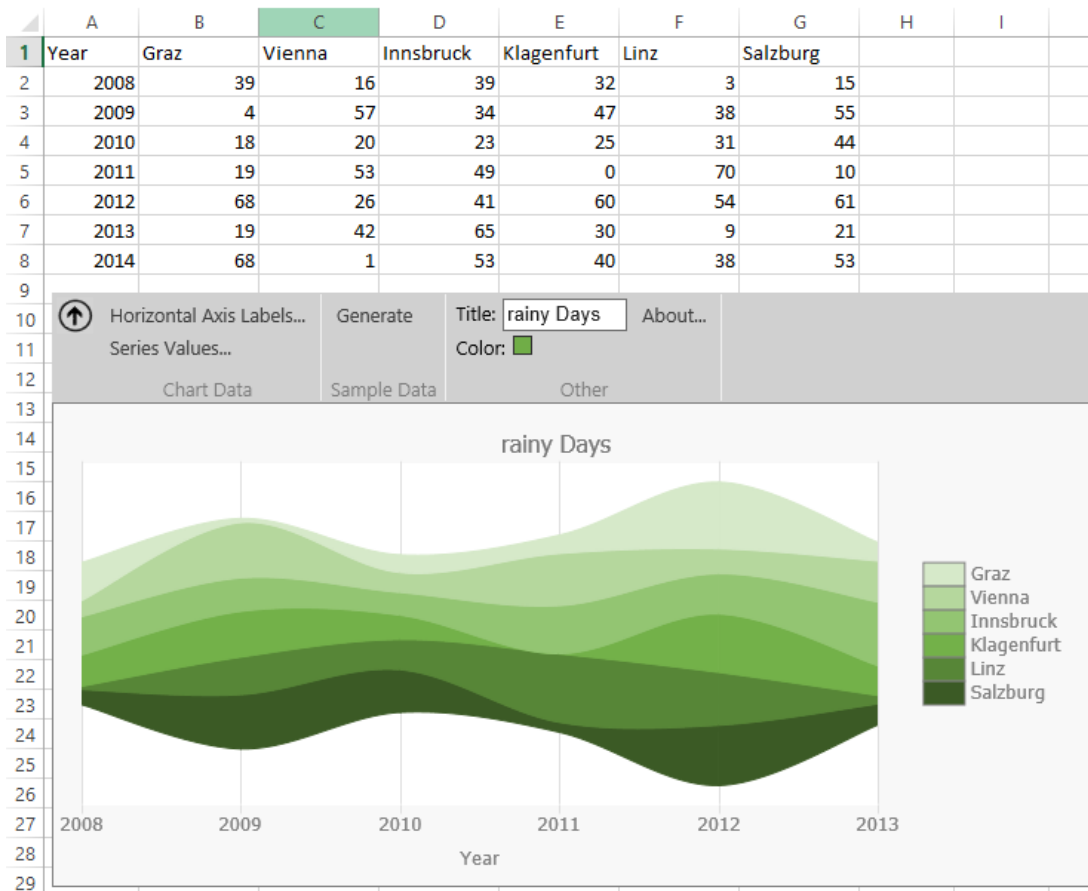

Figure 5.5: The Streamgraph app visualizes the number of rainy days in Austria in a period of six years. The colours are mapped to the columns that contain the values, the time data is displayed on the horizontal axis.
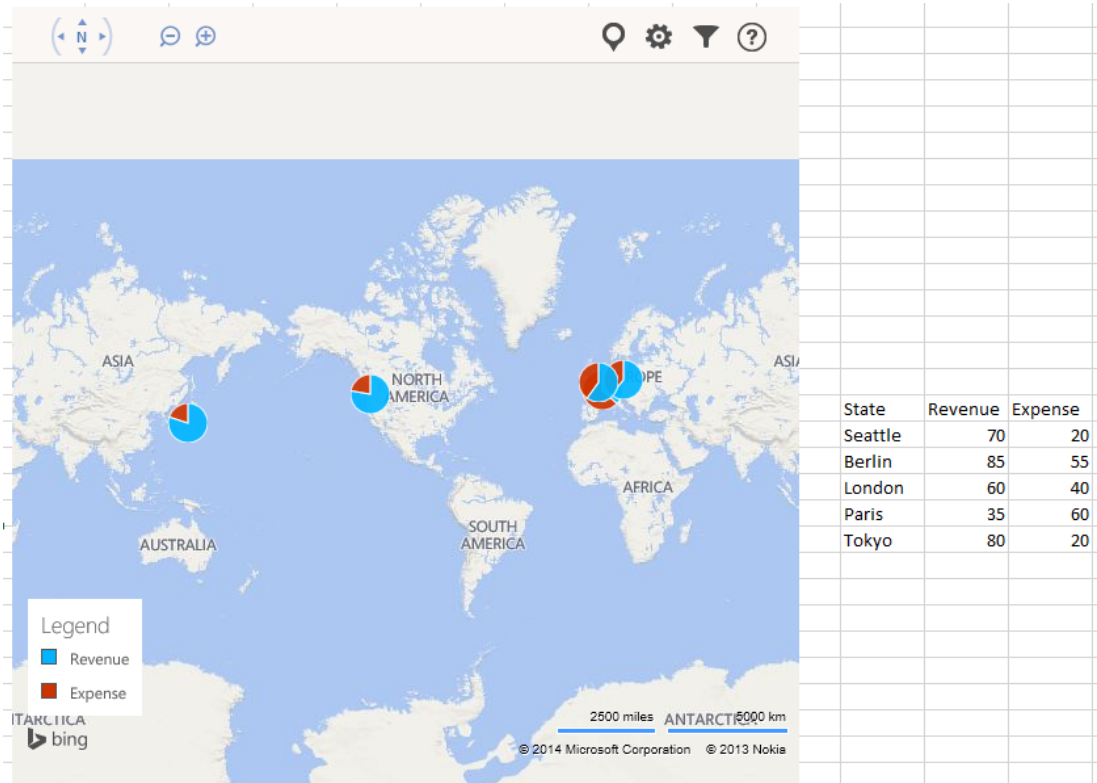
State	Revenue	Expense
Seattle	70	20
Berlin	85	55
London	60	40
Paris	35	60
Tokyo	80	20

Figure 5.6: The Bing Maps widget. The pie charts are shown in the map based on the provided sample data.
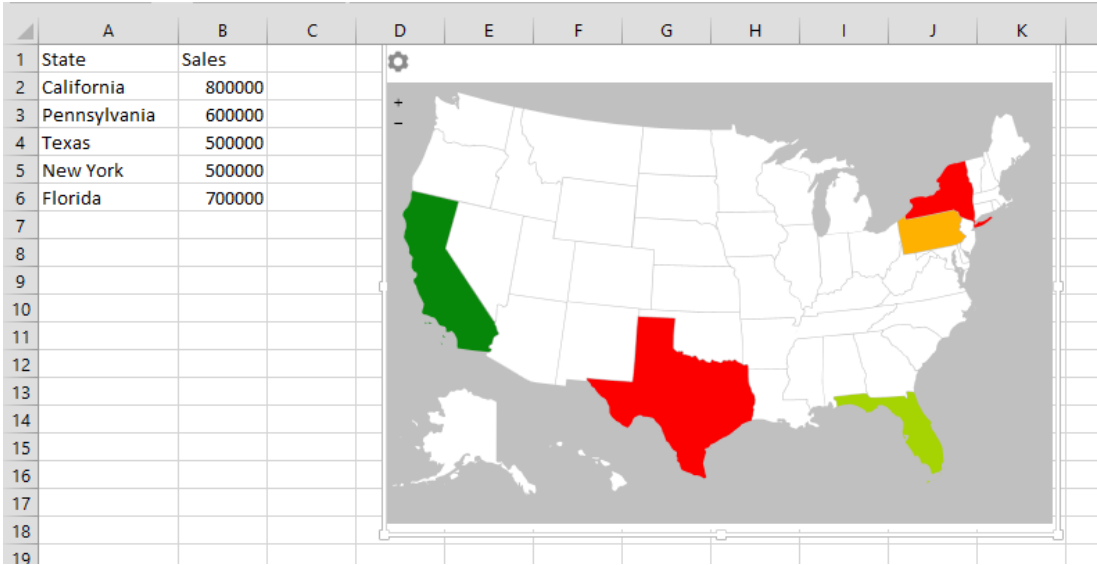

Figure 5.7: The Geographic Heat Map app. The numerical values are mapped to colours of countries in the map of the USA.

5.1.8 Gauge

Gauge [Revision Design, 2014] a free and very simple app that takes numerical value from a spreadsheet cell and displays it in a gauge based on user-defined minimum and maximum values. The users can choose between three different gauge looks.

5.2 Performance

Although the apps described in the previous section provide mostly simple functionality, some of them showed very poor performance. The SmartCharts app crashed all the time, despite the fact that the tests were done on the sample data. Some of the apps available in the Office Store did not work at all, such as the Radial Bar Chart [Patel, 2014b] or Slider Chart App [Zulu, 2014].

Chapter 6

Concluding Remarks

In the previous chapters, a short overview of visualisation possibilities with different applications and plug-ins for spreadsheet applications was given.

After direct comparison of visualisation possibilities, it can be easily concluded that the Microsoft Excel is clearly ahead of LibreOffice Calc. Although both spreadsheet applications have built-in visualization features, Microsoft Excel has a far better performance. Apart from that, a variety of plug-ins and apps is available for Microsoft Excel, whereas for LibreOffice Calc almost no additional visualisation features are available. Additionally, Microsoft Excel users can easily find a wide range of guides, tutorials and templates, both free and commercial, to not only create data visualisations, but also to present the data in an appropriate way, meaning the visualisation is as easy to interpret as possible, even with the maximum amount of data.

Such huge difference in the amount of available additional features could be explained by the fact that Microsoft Excel is a widely accepted and commonly used commercial application, suitable both for business as well as for private users. Although the explored plugins are powerful tools for visualisation of specific data types, most of the apps of the Microsoft Office Store turned out to be just fancy features for fast visualisations in special cases. It is clear that it is in Microsoft's interest to provide an appropriate range of features in order to be easily distinguishable from other spreadsheet applications, however a quantity should never overcome the quality.

As a free alternative, LibreOffice Calc is fully capable of visualising pre-processed data in standard chart types like bar charts or line charts. With some additional effort in fine-tuning, nice charts for smaller datasets can be easily done even in LibreOffice Calc. Nonetheless, in comparison to Microsoft Office it clearly has drawbacks regarding limitation in visualisation types and customisation and performance.

Bibliography

HI-CHART [2014]. *Chart-me*. 2014. `http://www.hi-chart.com/en/` (cited on page 19).

CodePlex [2014a]. *NodeXL: Network Overview, Discovery and Exploration*. 2014. `http://nodexl.codeplex.com/` (cited on page 13).

CodePlex [2014b]. *Social Network Importer for NodeXL*. 2014. `http://socialnetimporter.codeplex.com/` (cited on page 13).

devscope [2014]. *SmartCharts*. 2014. `http://www.devscope.net/products/SmartCharts` (cited on page 25).

Hansen, Derek, Ben Shneiderman, and Marc A. Smith [2010]. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2010. ISBN 0123822297, 9780123822291 (cited on pages 14, 18).

Harel, David and Yehuda Koren [2001]. "A Fast Multi-scale Method for Drawing Large Graphs". In: *Proceedings of the 8th International Symposium on Graph Drawing*. GD '00. London, UK, UK: Springer-Verlag, 2001, pages 183–196. ISBN 3-540-41554-8. `http://dl.acm.org/citation.cfm?id=647552.729397` (cited on page 14).

Microsoft Corporation [2010]. *Overview of PivotTable and PivotChart reports*. 2010. `http://office.microsoft.com/en-us/excel-help/overview-of-pivottable-and-pivotchart-reports-HP010342752.aspx?CTT=1` (cited on page 3).

Microsoft Corporation [2014a]. *Bing Maps*. 2014. `http://office.microsoft.com/en-us/store/bing-maps-WA102957661.aspx?queryid=63956a44-1cee-4ce6-85cb-0860931673eb&css=bing%20maps&CTT=1` (cited on page 27).

Microsoft Corporation [2014b]. *Buy Office 365*. 2014. `http://office.microsoft.com/de-at/business/compare-office-365-for-business-plans-FX102918419.aspx` (cited on page 3).

Microsoft Research [2014a]. *Histogram*. 2014. `http://office.microsoft.com/en-us/store/histogram-WA104042888.aspx` (cited on page 25).

Microsoft Research [2014b]. *Histogram 2D*. 2014. `http://office.microsoft.com/en-us/store/histogram-2d-WA104068025.aspx?queryid=b8bc194d-90a7-438c-a2f4-6512243939fa&css=histogram%202d&CTT=1` (cited on page 25).

Microsoft Research [2014c]. *Streamgraph*. 2014. `http://office.microsoft.com/en-us/store/streamgraph-WA104067192.aspx?queryid=cc20eb72-9a86-4465-a9cf-bfd031e49122&css=streamgraph&CTT=1` (cited on page 27).

Microsoft Research [2014d]. *Treemap*. 2014. `http://office.microsoft.com/en-us/store/treemap-WA104067194.aspx?queryid=0ff6e2e7-a4e7-4305-886b-794e6f587657&css=treemap&CTT=1` (cited on page 27).

Open Government Data Graz [2014]. *Population of Graz by year and age*. 2014. `http://data.graz.gv.at/daten/package/die-grazer-bev-lkerung-nach-bezirk-und-alter` (cited on page 8).

Patel, Keyur [2014a]. *Geographic Heat Map*. 2014. `http://office.microsoft.com/en-us/store/geographic-heat-map-WA103304320.aspx?queryid=7b450d15-61f3-4956-9e4e-f8eb70c92fa2&css=geographic%20heat%20map&CTT=1` (cited on page 27).

Patel, Keyur [2014b]. *Radial Bar Chart*. 2014. `http://office.microsoft.com/en-us/store/radial-bar-chart-WA103857593.aspx` (cited on page 30).

Revision Design [2014]. *Gauge*. 2014. `http://office.microsoft.com/en-us/store/gauge-WA103524919.aspx` (cited on page 30).

Social Media Reserach Foundation [2014a]. *About Social Media Research Foundation*. 2014. `http://www.smrfoundation.org/about-us/` (cited on page 13).

Social Media Reserach Foundation [2014b]. *NodeXL Graph Gallery*. 2014. `http://www.nodexlgraphgallery.org/` (cited on page 19).

Social Media Reserach Foundation [2014c]. *NodeXL Teaching Resources*. 2014. `https://nodexl.codeplex.com/wikipage?title=NodeXL%20Teaching%20Resources` (cited on pages 14–17, 19, 20).

The Document Foundation [2014a]. *LibreOffice*. 2014. `http://www.libreoffice.org/` (cited on page 7).

The Document Foundation [2014b]. *LibreOffice Extensions*. 2014. `http://extensions.libreoffice.org/extension-center/geooo` (cited on page 7).

The Document Foundation [2014c]. *The LibreOffice Calc Guide*. Chapter 3, pages 81-121. 2014. `https://wiki.documentfoundation.org/images/4/47/CG41-CalcGuideLO.pdf` (cited on page 9).

Twitter [2014]. *REST API Rate Limiting in v1.1*. 2014. `//https://dev.twitter.com/docs/rate-limiting/1.1` (cited on page 13).

Video2Brain [2014]. *Suchergebnisse für 'pivot'*. 2014. `https://www.video2brain.com/de/search.htm?search_entry=pivot` (cited on page 3).

Wikipedia [2014a]. *Force-directed Graph Drawing*. 2014. `http://en.wikipedia.org/wiki/Force-based_algorithms` (cited on page 14).

Wikipedia [2014b]. *Layered Graph Drawing*. 2014. `http://en.wikipedia.org/wiki/Layered_graph_drawing` (cited on page 14).

Zulu, Bongani [2014]. *Slider Chart App*. 2014. `http://office.microsoft.com/en-za/store/slidder-chart-WA104213984.aspx?redir=0` (cited on page 30).