STORYTELLING FROM SOCIAL DATA: DYNAMIC DATA EXPLORATION USING JMP

Volker Kraft
SAS Institute GmbH, JMP Division
Cologne, Germany
volker.kraft@jmp.com

Interacting with graphical displays and multivariate analysis tools helps to understand what the data are telling us. The interactivity of JMP for all types of data, including social, geographic and time-series data, helps with efficient visualization and modeling, supports decision-making from data, and facilitates the communication of findings and results.
In this hands-on workshop we used recent social data to illustrate the power of dynamic tools for data exploration, visualization and analysis, and explored the role of interactivity in statistical education. Participants were asked to download and install the 30-day trial version of JMP (www.jmp.com/trial for Windows or Mac) before the workshop.

INTRODUCTION

This paper recaps what was done during the practical JMP Workshop, which was offered to attendees of the IASE Roundtable Conference 2016 on July 21st and 22nd 2016 in Berlin.

The following section describes the workshop goal and the software tools used by the participants. In case you want to run the analyses yourself (again), just download the JMP Free Trial (2016) and the Workshop Content (2016). During the workshop JMP (Pro) 12 was used, but all steps can also be carried out in later versions of JMP or JMP Pro. The content includes the JMP journal (Fig. 1) with links to the data sets and to more information online. All data sets have stored analysis scripts, so that all results can be reproduced easily by a single click. The figures at the end of this paper show some sample analyses.

In this paper the three cases using diverse social data are summarized briefly. For each case we outline the questions and the tools used to find answers. Please note there is no single "standard solution", therefore we recommend to download the software and workshop content for your own hands-on. Feel free to play around and learn from the examples, before you start to discover the stories told by your own data.

Our conclusions discuss some feedback from the workshop and what this means for teaching settings.

WORKSHOP

Goal

The goal of each 90-minute workshop was to let attendees experience themselves how engaging, enjoyable and efficient the analysis of social data can be nowadays. It was proven that intuitive and interactive software tools can make multivariate analysis easily accessible. A graphical point-and-click user interface keeps the user in a "data-driven" mode, leading to interesting discoveries without the need to involve a trained statistician or programmer. Visual summaries and graphs not only yield answers directly (and probably more questions), which support better decision-making, but also turn the process of data import, management, analysis and reporting into an exciting journey.

Tools

The best environment for this experience is "JMP in action", which combines a curious user with access to interesting data and JMP software. Probably you have lots of interesting data sources in mind, but if you don't have an idea about the JMP look and feel yet, you are invited to watch *JMP in 60 Seconds* (2016).

Most use cases in the workshop were in the areas of data visualization, multivariate analysis and (exploratory) modeling. Other JMP applications like experimental design or quality management/Six Sigma, as well as custom applications with direct interfaces to SAS, R, MATLAB and Excel, were just mentioned but not covered during the hands-on portion of the workshop. For more information, please refer to JMP Application Areas (2016).

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. iase-web.org/Conference_Proceedings.php

As expected, workshop attendees had different levels of JMP experience, from first-time user to power-user. In case you are new to JMP and want to learn some basics, the journal links some helpful resources in the section "*Want to See More?"*. Also see Academic Resources for Teaching and Learning (2016) or search for and access the latest content in the JMP Academic Community (2016).

Cases

Three cases were selected to tackle different kinds of questions: The well-known Titanic Passenger data set provided information about 1309 passengers who were on the Titanic, including the survival status of each passenger. The second case showed the dynamic import of more recent online data: Results from polls carried out before the Brexit referendum. The detailed referendum results from different regions of the UK revealed some interesting insights. Finally, data from the recent refugee crisis in Europe was visualized geographically.

For each case the following chapters summarize the data source, the problem, JMP tools used during the analysis and some findings. More detailed instructions can be found in the JMP journal, which can be downloaded at Workshop Content (2016). Analysis results can be reproduced and modified by running the table scripts.

CASE 1: TITANIC (EXPLORATORY MODELING)

Data: *Titanic Passengers.jmp* (original version in the JMP Sample Data Library, slightly modified, scripts added)

Problem: Was there an equal chance to survive the Titanic?

JMP tools:

- *Distribution*, with *Data Filter*
- *Fit Y By X* to explore bivariate relationships
- *Graph Builder* for data visualization
- *Partition* to build a classification tree
- *Fit Model* to build a logistic regression model, with *Stepwise* regression for variable selection
- *Profiler* to explore the modeled relationships

Findings: Exploratory modeling revealed 'Sex', 'Age' and 'Passenger Class' as the most important factors explaining 'Survived'. The Profiler helped to interpret the model. By Saving the Prediction Formula to the data table the chance to survive the Titanic disaster could be estimated for other passengers.

CASE 2: BREXIT (GRAPHING)

a) *Brexit Polls (Bubble Plot)*

Data: Online from Brexit poll tracker (2016)

Problem: Surprised, or should we have known (from polls)?

JMP tools:

- *Internet Open* to load HTML table data into JMP
- Variable post-processing (see journal script)
- *Stack* and *Recode*
- *Bubble Plot*, with export as *Interactive HTML*

Findings: The time series data from polls about "Remain or Leave?" was visualized by an animated Bubble Plot, showing the dead even close to the referendum date. The animated plot was exported as interactive HTML.

b) *Brexit Results (Graph Builder)*

Data: Referendum results from BBC News (2016)

Problem: Variation in the vote across the UK? How decisive were the actual results?

JMP tools:

- *Tabulate* to summarize data
- *Graph Builder* to create sorted bar charts
- Row states *Hide & Exclude* (Gibraltar)
- Distribution > *Normal Fit*, Save Density Formulas

- *Graph Builder* to visualize normal density curves by Region 2

Findings: While regions like London and Scotland were more on the remain side (Leave < 50%), the English Midlands were clearly identified as being in favor to leave.

CASE 3: REFUGEE CRISIS (DATA FILTER, MAPPING)

Data: First time asylum applicants by citizenship, age and sex. Monthly data (rounded). Source = EUROSTAT 7-2016. Count = # Persons.

Problem: How has the number of first time asylum applicants changed over the last 2,5 years? Which countries received the most applicants from ext. EU, Syria or Nigeria?

JMP tools:

- *Distribution*
- *Graph Builder* with *Local Data Filter*
- *Mapping* using shape files and background maps

Findings: Asylum applicant statistics from EUROSTAT could be analyzed with help of the Data Filter to exclude e.g. total numbers. Geographic information was shown by Graph Builder in various ways, including wrapping, overlay and mapping. Trends over time were recognized and explained and differences between countries were pointed out.

WANT TO SEE MORE?

During the workshop several resources were recommended, which can be used for self-learning "to see more" examples and analytical platforms in action. Most of this content, but especially the two "libraries" below, are also embraced by instructors worldwide in teaching their classes.

Here are the best starting points:

- JMP Application Area: Data Visualization and Exploratory Data Analysis with JMP (2016)
- JMP Help menu: *JMP Help* and *Books* for documentation. *Sample Data* offer around 500 datasets, most with stored scripts.
- JMP Learning Library (2016) and JMP Case Study Library (2016)
- JMP Webcasts On-demand (2016)
- JMP e-Learning (2016)
- Book: Goos, P. & Meintrup, D. (2015)

For more specific resources you may want to check out the JMP Academic Community (2016), JMP on YouTube (2016) or just use your favorite search engine, e.g. Google.

CONCLUSION

During these two workshops we showed how easy it is to get data into JMP, e.g. by using sample data, importing table data from a web site or opening Excel data. Three example cases were selected, and we focused on exploratory data analysis to understand the data and explanatory analysis with the goal of story-telling.

All hands-on exercises were carried out successfully by attendees who already teach with JMP, but also by attendees with no prior JMP experience. Furthermore, using both Windows and Mac versions of JMP at the same time did not cause any problems.

All received feedback from attendees confirmed that the data exploration and analysis in JMP was easy to follow, and that the created visuals were great for telling the story of the data. The participants' feedback can be summarized from two perspectives:

From a story-telling perspective, the audience caught fire seeing the examples of data visualization and exploratory data analysis (EDA) tools in JMP. It became obvious that in JMP dynamic graphics go hand-in-hand with statistics and analytical tools, a synergy which helps you to interactively share findings in ways decision makers can readily act upon.

From a teaching perspective, the following JMP advantages could be demonstrated:

- The nature of the JMP user interface allows to focus on data, problems and concepts, rather than teaching tools and commands
- The dynamics and visualizations of JMP are best-in-class to engage students

- Using JMP in teaching is fun, leveraging the existing sample data, tutorials and teaching materials, e.g. the One-Page Guides from the JMP Learning Library (2016)

The JMP Blog (2014) *"Fun and effective: Teaching statistics with JMP"* mentions some quotes from noteworthy educators worldwide about why they like teaching with JMP.

REFERENCES

Academic Resources for Teaching and Learning (2016). JMP Website. *http://www.jmp.com/teach*

BBC News (2016). Referendum results. *http://www.bbc.com/news/politics/eu_referendum/results*

Brexit poll tracker (2016). Financial Times website. *https://ig.ft.com/sites/brexit-polling/*

Goos, P. & Meintrup, D. (2015). *Statistics with JMP: Graphs, Descriptive Statistics and Probability*. Chichester, UK: Wiley.

JMP Academic Community (2016). Online community for academic users in JMP User Community. *http://www.jmp.com/jac*

JMP Application Area: Data Visualization and Exploratory Data Analysis with JMP (2016). JMP Website. *http://www.jmp.com/dataviz*

JMP Application Areas (2016). JMP Website. *http://www.jmp.com/applications*

JMP Blog (2014). *Teaching statistics with JMP.* *http://blogs.sas.com/content/jmp/2014/06/16/fun-and-effective-teaching-statistics-with-jmp/*

JMP Case Study Library (2016). JMP Website. *http://www.jmp.com/cases*

JMP e-Learning (2016). *JMP Software: A Case Study Approach to Data Exploration*. JMP Website. *http://www.jmp.com/en_us/academic/jmp_ecourses.html*

JMP Free Trial (2016). 30-days trial of recent JMP version for Windows or Mac. Cary, North Carolina: SAS Institute. *www.jmp.com/trial*

JMP in 60 Seconds (2016). YouTube. *http://www.jmp.com/jmpin60*

JMP Learning Library (2016). JMP Website. *http://www.jmp.com/learn*

JMP on YouTube (2016). YouTube channel *JMPSoftwareFromSAS*. *https://www.youtube.com/channel/UCPOQfD32PLMVdTg-38OwjCA*

JMP Screenshot Gallery (2016). JMP Website. *http://www.jmp.com/en_us/academic/screenshot-gallery.html*

JMP Webcasts On-demand (2016). JMP Website. *http://www.jmp.com/ondemand*

Workshop Content (2016). JMP journal and data sets with stored scripts for download. *https://community.jmp.com/docs/DOC-10051*

FIGURES

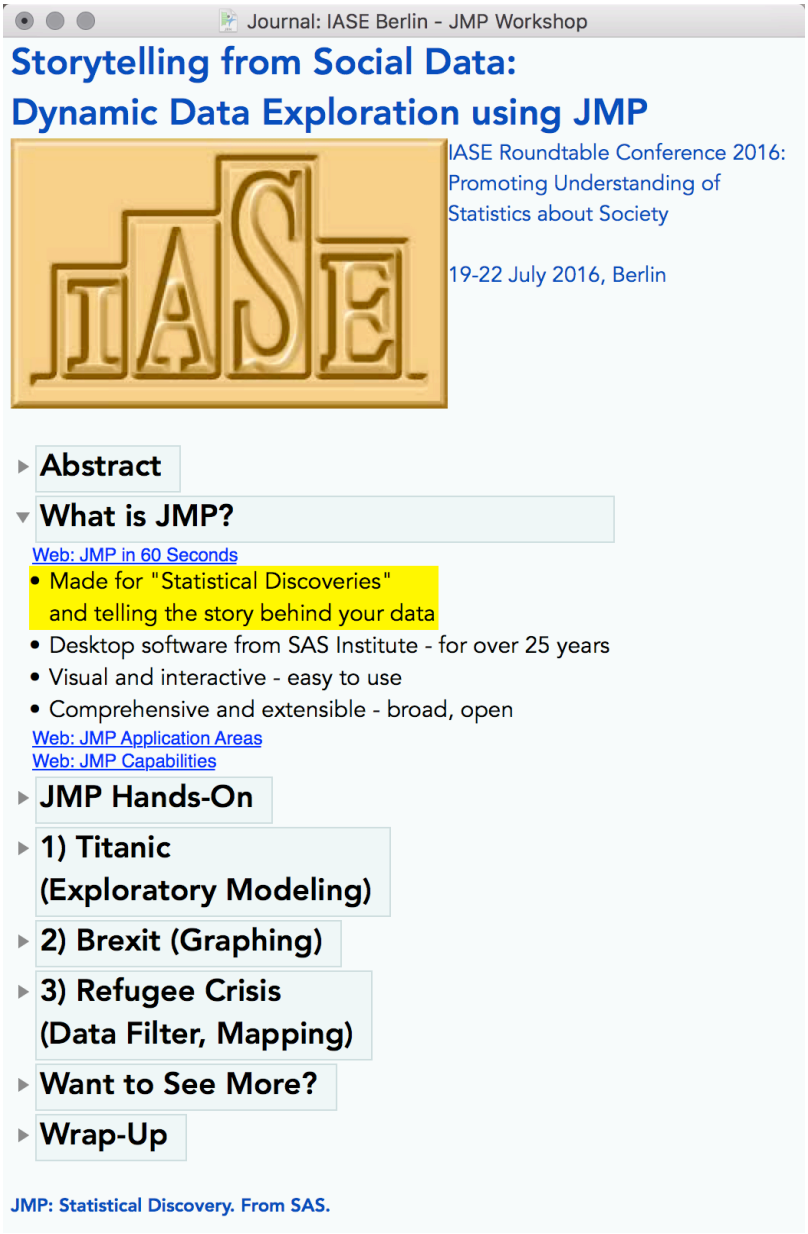

Figure 1. Overview of the JMP Journal

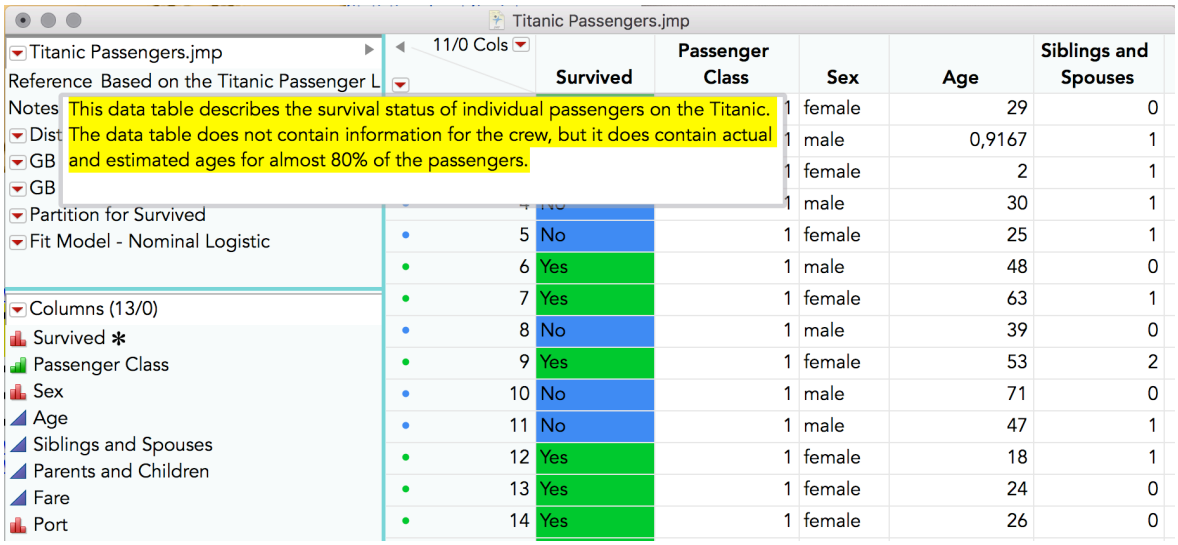

Figure 2. Part of JMP data table "Titanic Passengers.jmp": Table notes (in yellow), stored table scripts (with red triangles), column panel and data (subset)

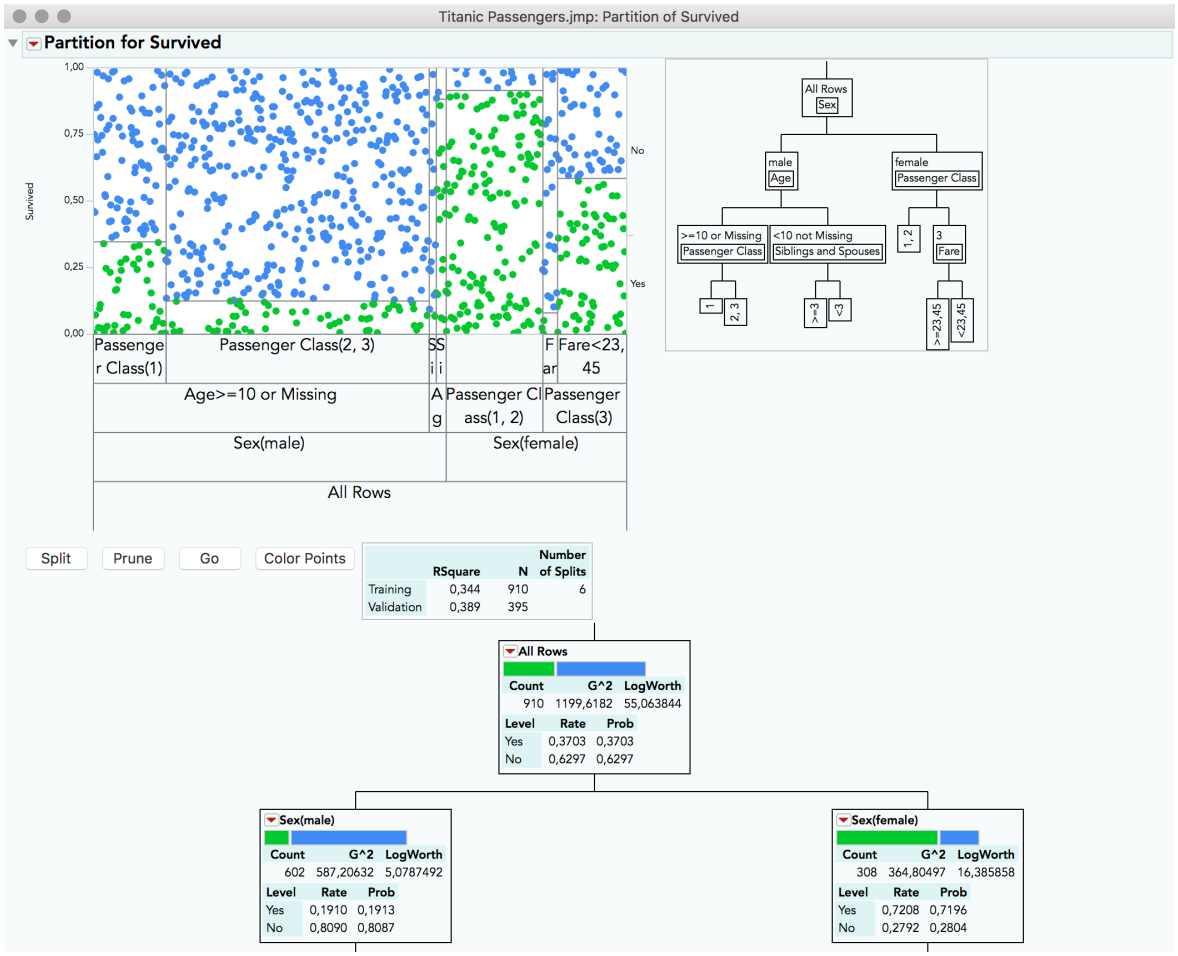

Figure 3. Case 1 – Partition for Survived (extract)

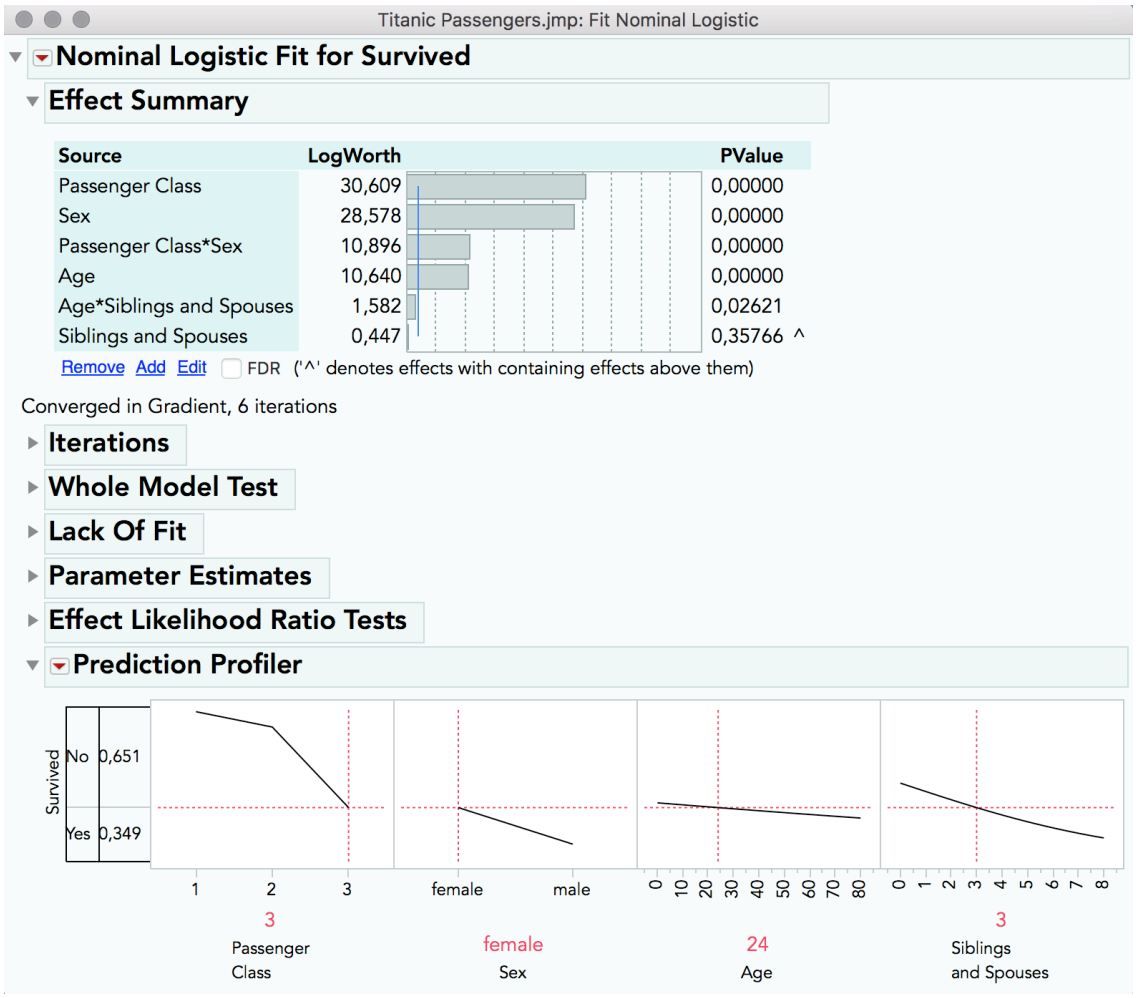

Figure 4. Case 1 – Nominal Logistic Fit for Survived

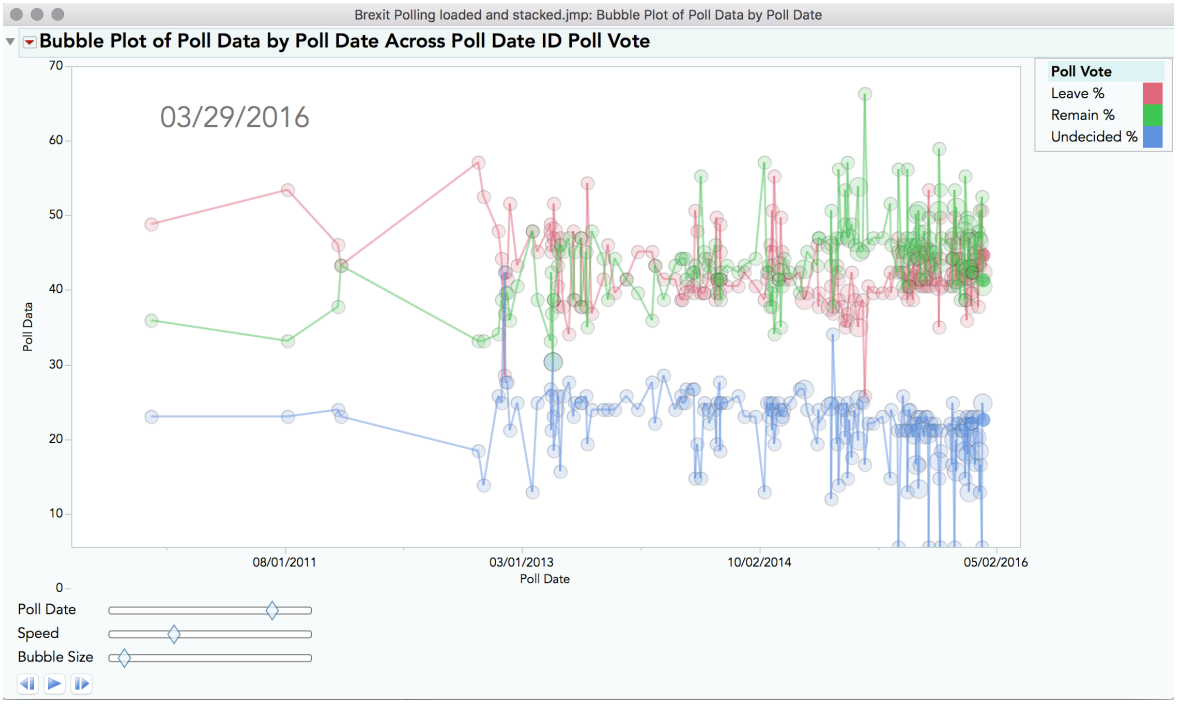

Figure 5. Case 2a – Bubble Plot of Poll Data (animated graph)

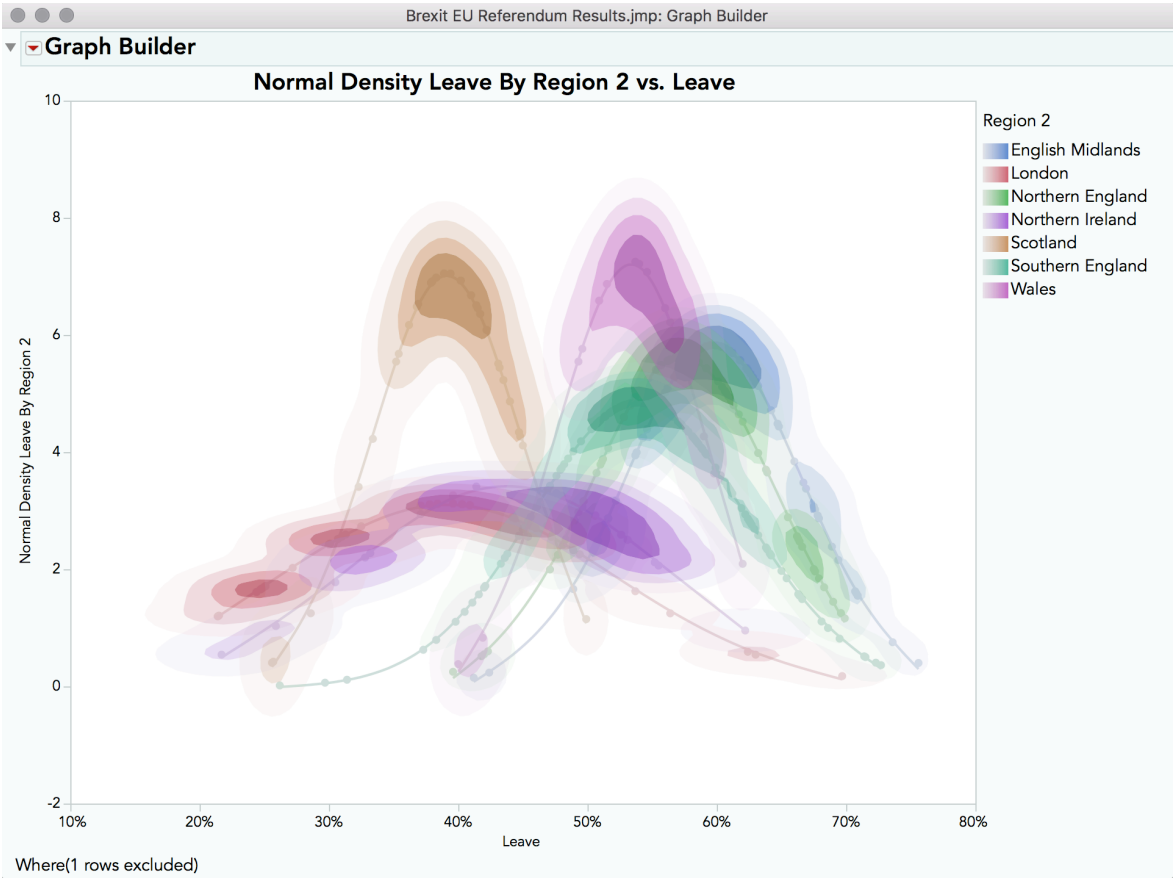

Figure 6. Case 2b – Normal Densities for Leave By Region 2

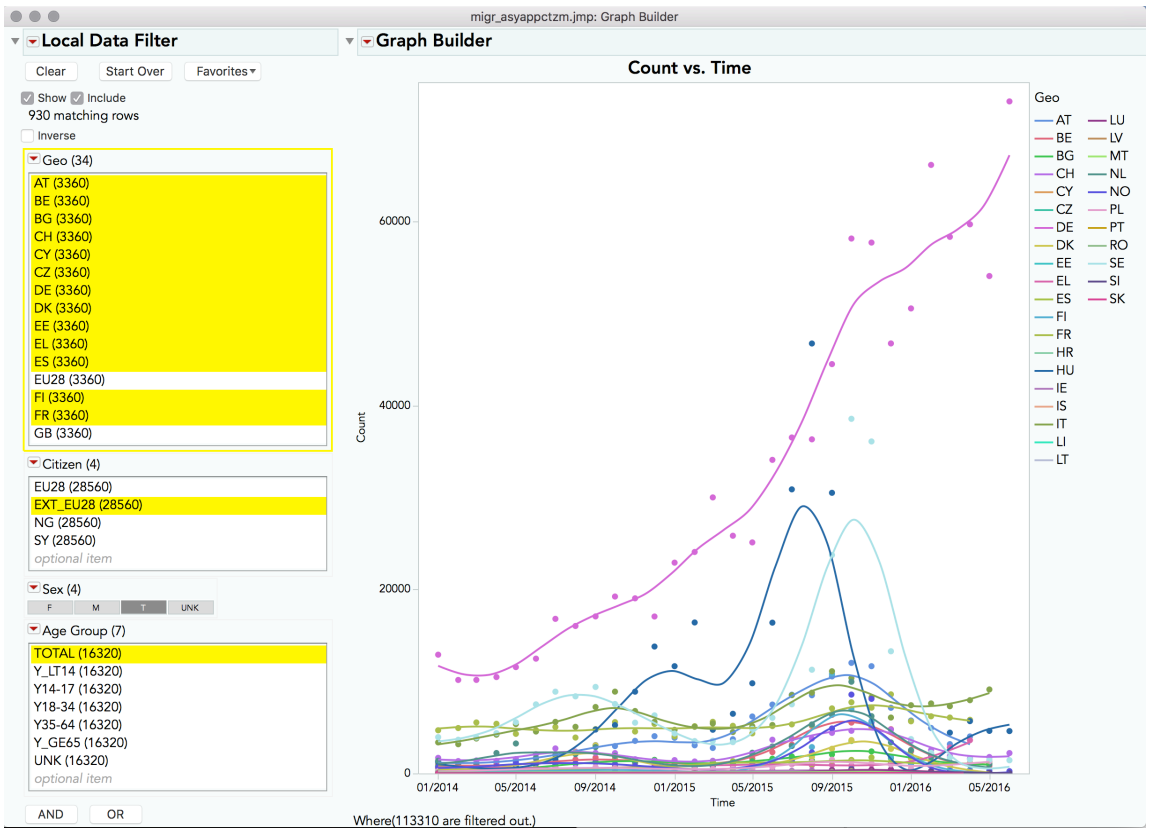

Figure 7. Case 3 – Count vs Time by Country, with Data Filter

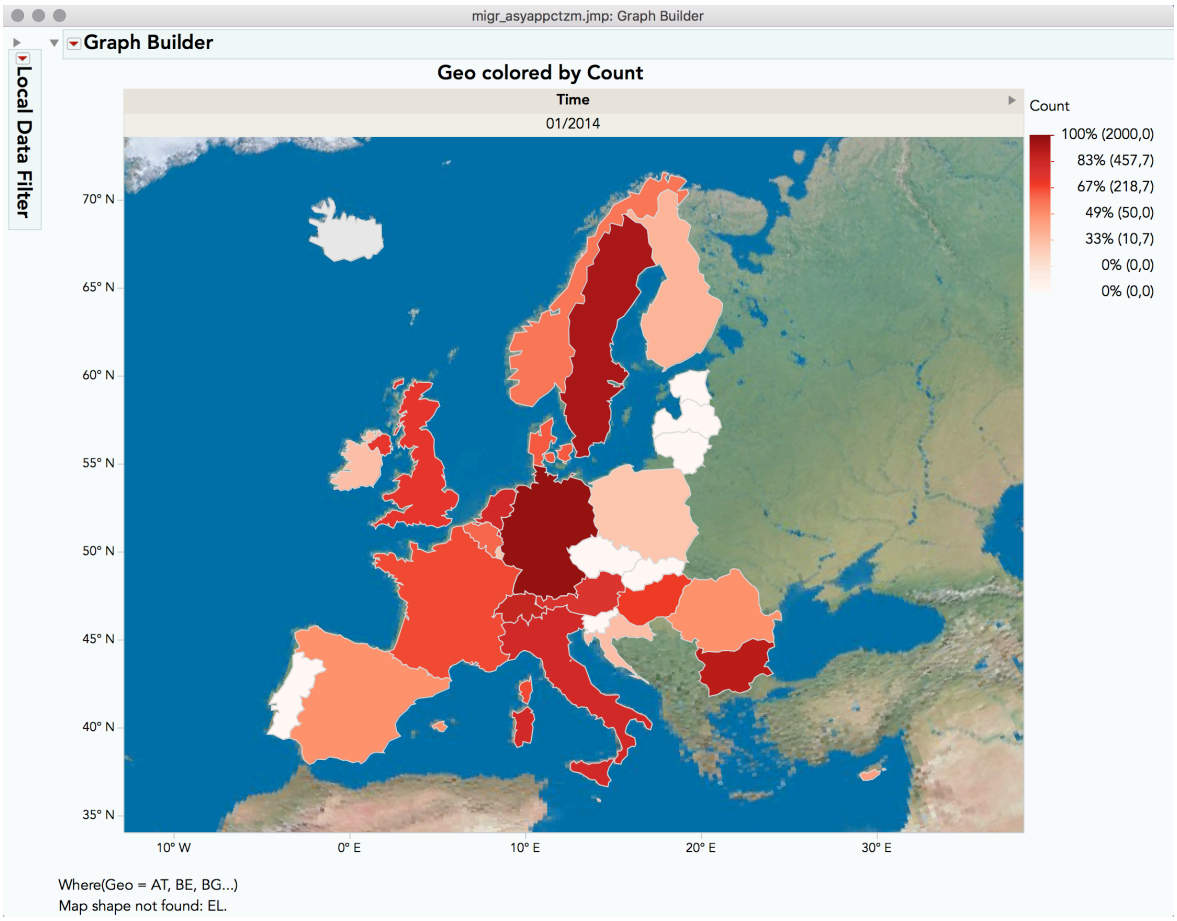

Figure 8. Case 3 – Asylum applicants by country