IN4086-14 Data Visualization
Final Project Report

Federico Falconieri 4519892 & Xiaoxu Gao 4504348

January 27, 2016

1. PROBLEM AND MOTIVATION: WORLD TOURISM

Many data are available regarding world wide tourism: absolute number of outbound and inbound travellers per country and per year, amount of money spent in a country by foreign tourists, total amount of money spent in other countries by one country's outbound tourists. To derive meaningful information by looking at the sheer numbers is challenging though, and only trivial conclusions can be deduced. In order to understand these numbers in a more meaningful and interesting way visualization is needed. InfoVis allows easily to make comparison between countries, to see the global and local trends in a set period of time. Furthermore visualization allows to study the correlation between tourism and economics indicator like the PPP*.

2. DATASET

Three datasets have been used for this project:

- For the tourism information (inbound and outbound tourism, inbound and outbound tourism expenses) `http://data.un.org/DocumentData.aspx?id=372`

- For the population `http://esa.un.org/unpd/wpp/DVD/Files/1_Indicators%20(Standard)/EXCEL_FILES/1_Population/WPP2015_POP_F01_1_TOTAL_POPULATION_BOTH_SEXES.XLS`

- For the PPP: `http://data.un.org/Data.aspx?d=WDI&f=Indicator_Code%3ANY.GNP.PCAP.CD`

Since the smallest subset of years that these three datasets had in common was 1995-2014 this interval was used.

3. VISUALIZATIONS AND RESULTS

A total of 8 visualizations (v1, v2,..., v8) has been produced for this project. V1 and v2 are line charts, v3 is a combination of a line chart and a bar chart, v4 and v5 are steamgraphs, v6 is a PCP, v7 and v8 are cloropleth. The couples (v1,v2), (v4,v5), (v7,v8) are the same visualization with different data (inbound tourism and outbound tourism), this is a design choice to make comparison between the two versions easier (eyes beat memory).

3.1 Line charts (v1,v2)

Figure 1 shows the trend of the inbound tourism in the years 1995-2014 for China, France and Italy. Up to three countries can be interactively selected from a drop down list and plotted together. Three was selected after a number of trials as it is a reasonable number to prevent cluttering. There are three attributes: country (categorical), inbound/outbound tourism (quantitative), year (ordinal). This visualization, line chart, was chosen instead of others since the main task was to find and show the trend in the years 1995-2014. Being the key attribute "year" ordered, a line chart was preferred to a bar chart. The marks used are points and line connecting the points. Lines connect points with the same "country" categorical attribute. This way a separate trend is shown for each country and comparison are possible. Vertical position is the channel used for the quantitative attribute "inbound/outbound tourism", and the horizontal is used for the ordinal attribute "year". Hue is the color channel used for the categorical attribute "country". According to Mackinlay[1], position is the best channel to visualize quantitative data, whereas hue and position are the two best channels for categorical data. Hue and separation are also a good choice since they are highly separable. Each point is surrounded by a circle of variable radius that represents through the channel area the attribute absolute inbound/outbound expenses (respectively inbound expenses for v1 and outbound expenses for v2). For example the inbound absolute expenses are the money spent by a country's inbound tourists during their stay. This value is expressed in millions of US $.

---

*PPP is the Gross Domestic Product, at purchasing power parity, per capita.

3.2 Line + bar chart (v3)

In Figure 2 and 3 two separate visualizations are shown together in order to expose their correlation. Data regarding one country at a time are visualized. Country can be changed interactively. Absolute outbound tourism data are visualized with a bar chart, the trend of PPP is visualized with a line chart. Hue is used to differentiate between the two charts. The main task of this visualization was to show the trend of PPP and its effect year by year on the outbound tourism. Line chart are better for trends and bar charts are well suited for comparison of quantitative data (in this case the comparison of the quantitative attribute "outbound tourism"among the ordinal attribute "year"). Points and lines are the marks for the PPP, length is the mark for outbound tourism. These choice were made accordingly to Mackinlay[1]'s research. Channels used for the PPP: vertical position for the quantitative attribute "money", horizontal for the ordinal attribute "year". Channels used for outbound tourism: vertical length. If the two charts are considered two separate categories then, always according to Mackinlay[1] the color channel hue is an effective way to separate categorical data.

3.3 Steamgraphs (v4, v5)

Figure 4 and 5 are steamgraphs of respectively the inbound and the outbound tourism of the G20 countries in the period of years 1995-2013. The main task of this visualization was to show the global trend, the secondary to show each country's trend to compare them. It's secondary because individual countries trend can be already visualized (and in a better way) in sections 3.1 and 3.2. Data have three attributes: country (categorical), year (ordered), inbound/outbound tourism (quantitative); the steamgraph is a good way to visualize data with these characteristics. The marks used are points, lines connecting the points of every country (exactly as for 3.1), and mainly area. The channels used are color hue for the country (categorical) and vertical height for the quantitative attribute. Both good choices according to Mackinlay[1]. One could argue that area is not the best attribute to show quantitative attributes, but it must be kept in mind that this is used as a secondary channel to discriminate among countries (the primary is hue) and discrimination among the various countries is only the secondary task of this visualization. The main one, to show the G20 tourism trend, is accomplished well by the steamgraph since the top part of the chart behaves exactly like a simple line chart would do.

3.4 PCP (v6)

Figure 6 shows a Parallel Coordinate Plot. Data have 1 key categorical attribute, country, and 5 quantitative attributes: PPP, inbound tourism, outbound tourism, inbound expenses per tourist, outbound expenses per tourist. All country are showed at each time, only one year is visualized at each time. The task of this visualization was to study the correlation between these 5 attributes in the same visualization (multidimensional data), therefore the PCP was a good choice. Furthermore, axis (attributes) can be rearranged interactively, so correlation can be studied for every possible couple of attributes. The marks used are lines. Each line represent a data point, connecting all the values with the same categorical attribute (country). The channel used is the vertical position on the axis for quantity and this, as stressed before, follows Mackinlay[1]'s ranking. Color is used for focus: when a portion of an axis is selected all the lines outside the selection become grey and with low saturation. This way single data points can be isolated across all the attribute (it's a form of pop out). See Figure 7 for an example.

3.5 Cloropleths (v7, v8)

Figures 8 and 9 show a cloropleth. Data are the same used for the PCP, one year is selected at a time and it can be changed interactively with a slider. The task for this visualization was to make a geographical comparison between the countries' tourism data immediate. v7 shows inbound tourism and v8 outbound tourism. Inbound/outbound tourism is shown through hue, data regarding expenses are expressed through interactive labels that appear when hovering a country. Cloropleths can be zoomed in to visualize better small countries. The mark used is the area of each country. The visual channel used is color saturation for the quantitative attribute inbound and outbound tourism. Countries without data were left grey.
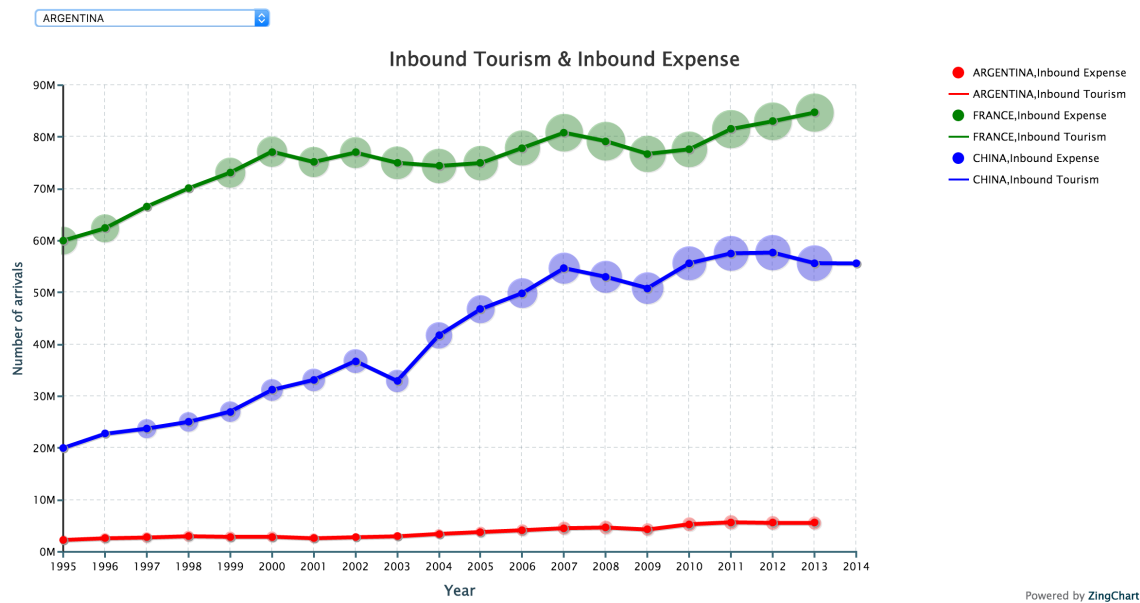
Appendices

Graphics

Figure 1. Line chart. Absolute number of arrivals for China, France, Poland and Italy between 1995 and 2014.
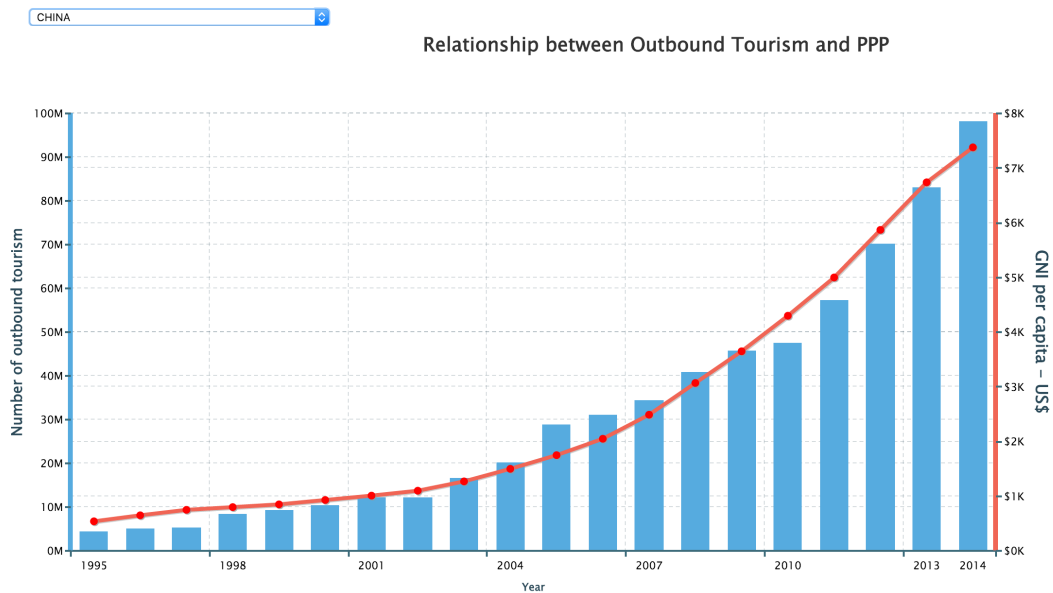
Figure 2. Lines and bars chart of Chinese outbound tourism and PPP. Correlation between PPP (red line) and outbound tourism (blue bars) is highly visible. China went from 5 millions of outbound tourists in 1995 to roughly 100 millions outbound tourists in 2014.
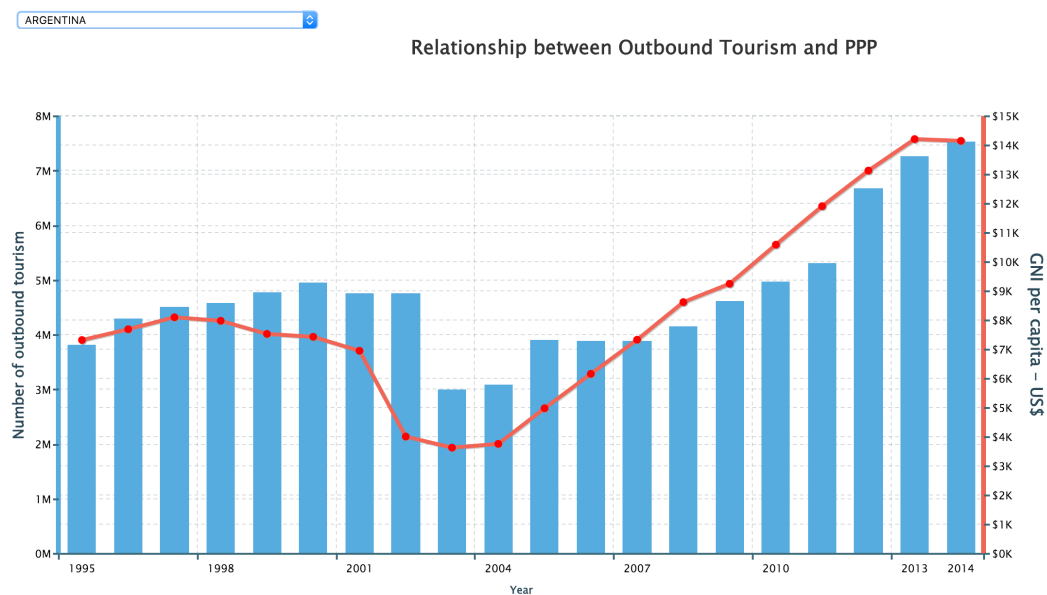

Figure 3. Lines and bars chart of Argentinian outbound tourism and PPP. As for Figure 2 correlation between PPP (red line) and outbound tourism (blue bars) is highly visible. 1998-2002 Argentine great depression[2] and its later effects on the country's economy are here clearly visible, both in terms of PPP and outbound tourism.
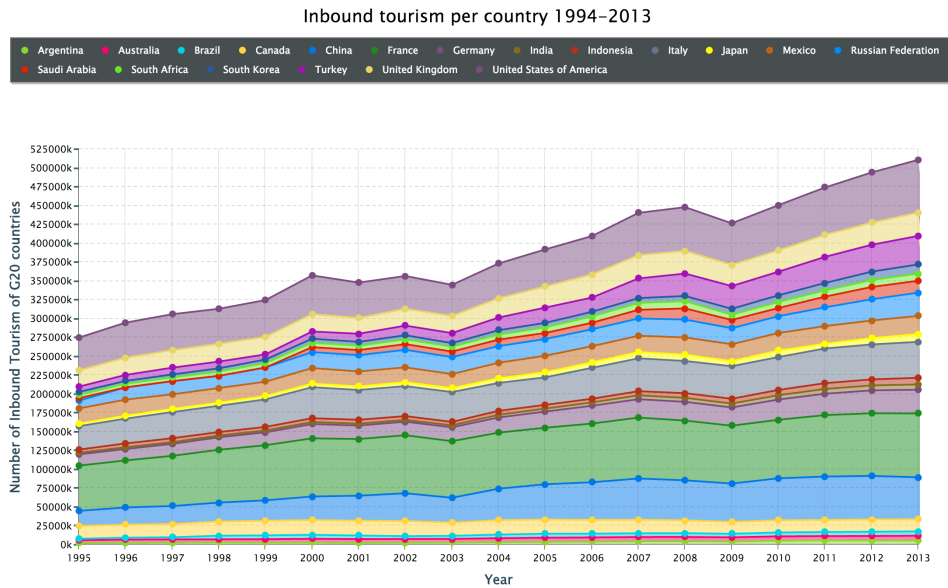
Figure 4. Steamgraph of inbound tourism of G20 economies (minus the EU) in the period 1995-2013. The 2008 global great recession[3] is clearly visible as well as the early 2000s recession [4] (enhanced in the US especially by the Dot-com bubble[5]).
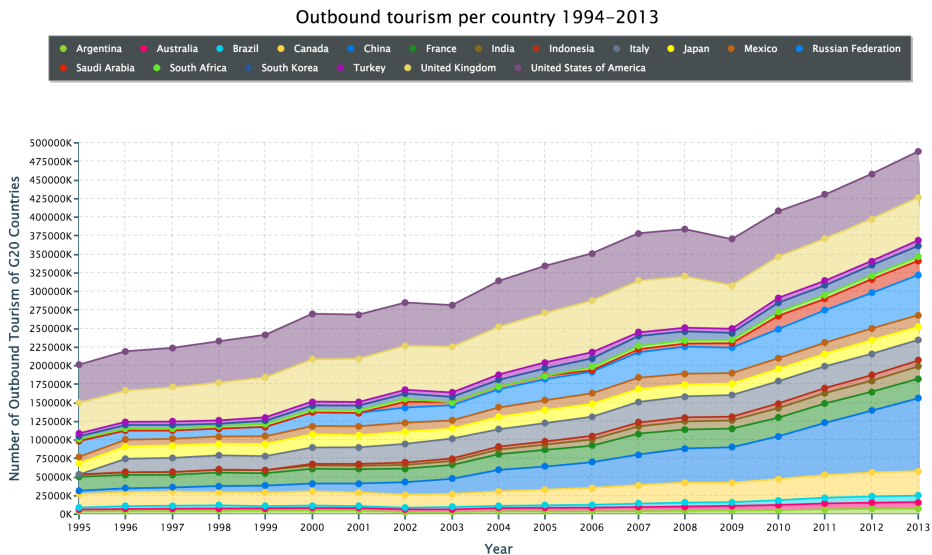

Figure 5. Steamgraph of outbound tourism of G20 economies (minus the EU and Germany, for which data were not reliable) in the period 1995-2013. Similarly to Figure4 the 2008 global great recession[3] is clearly visible. This graph shows also the remarkable growth in outbound tourism of the Russian Federation (in light blue, positioned centrally) and most of all of China (darker blue, positioned in the lower part of the graph). The growth of the B.R.I.C.S.(the five major emerging national economies: Brazil, Russia, India, China and South Africa) makes the effect of the early 2000s recession[4] less visible; by looking carefully, though, the United States outbound tourism shrank of almost 1/3 between 2000 and 2003.
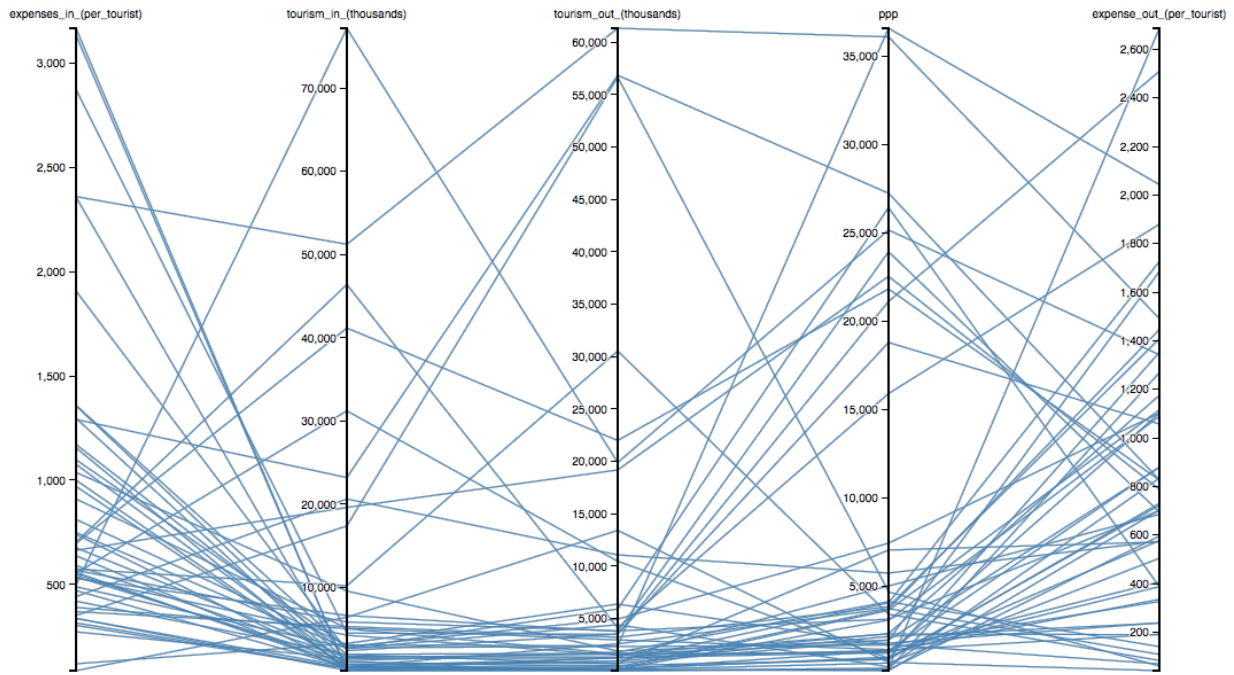
Figure 6. Parallel Coordinate Plot of the following attributes: absolute inbound tourists, absolute outbound tourists, PPP, average expenses of inbound tourists (in US$), average expenses of outbound tourists(in US$). One year is visualized at each time, in this case the year 2000. The order of the attributes can be moved in real time. The configuration showed provides a way to study correlation between PPP and outbound tourism and expenses. As expected there seems to be a correlation between outbound tourism and PPP, as well as PPP and outbound expenses. Countries with an average PPP above 15000 (US$) show a direct correlation with outbound expenses. Countries with a low average PPP, instead, show all sort of outbound expenses. We hypothesise that this is due to the fact that some countries with a low PPP usually have a lot of inequality in the way money are distributed among the population. Therefore the PPP, being an average value, is misleading. The few people that can travel outside these poor countries are extremely wealthy and consequently able to spend a lot when they are abroad.
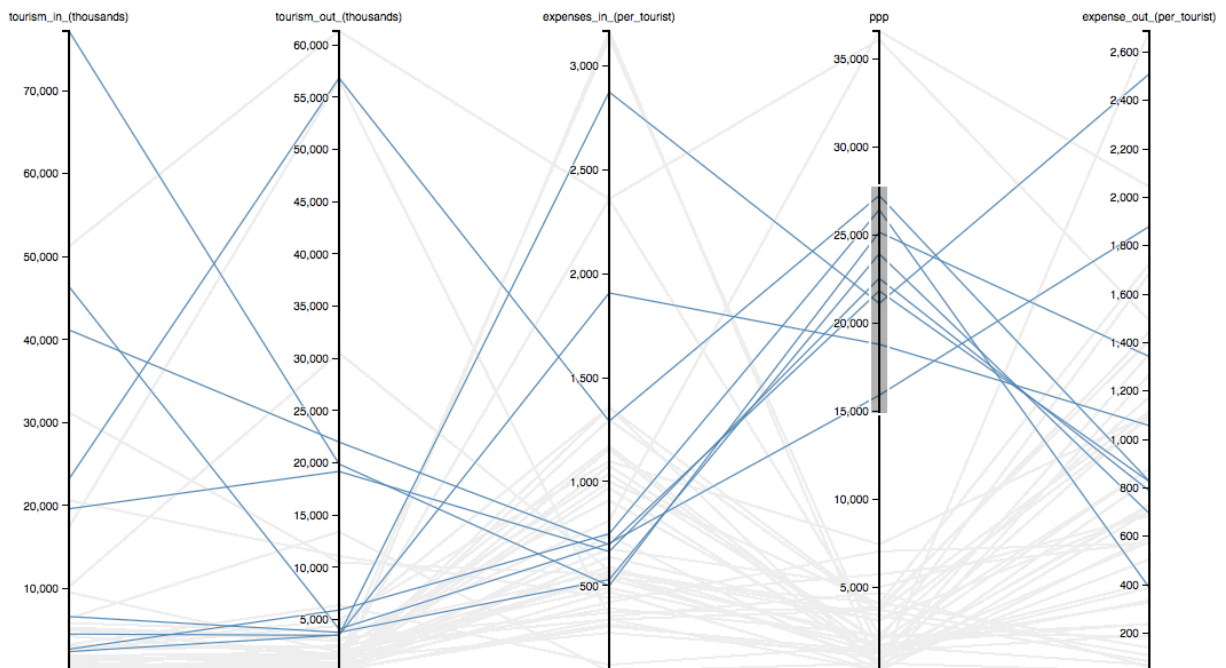

Figure 7. Same PCP as in Figure 6, only a portion of the attribute PPP is selected and therefore highlighted by making the other lines grey
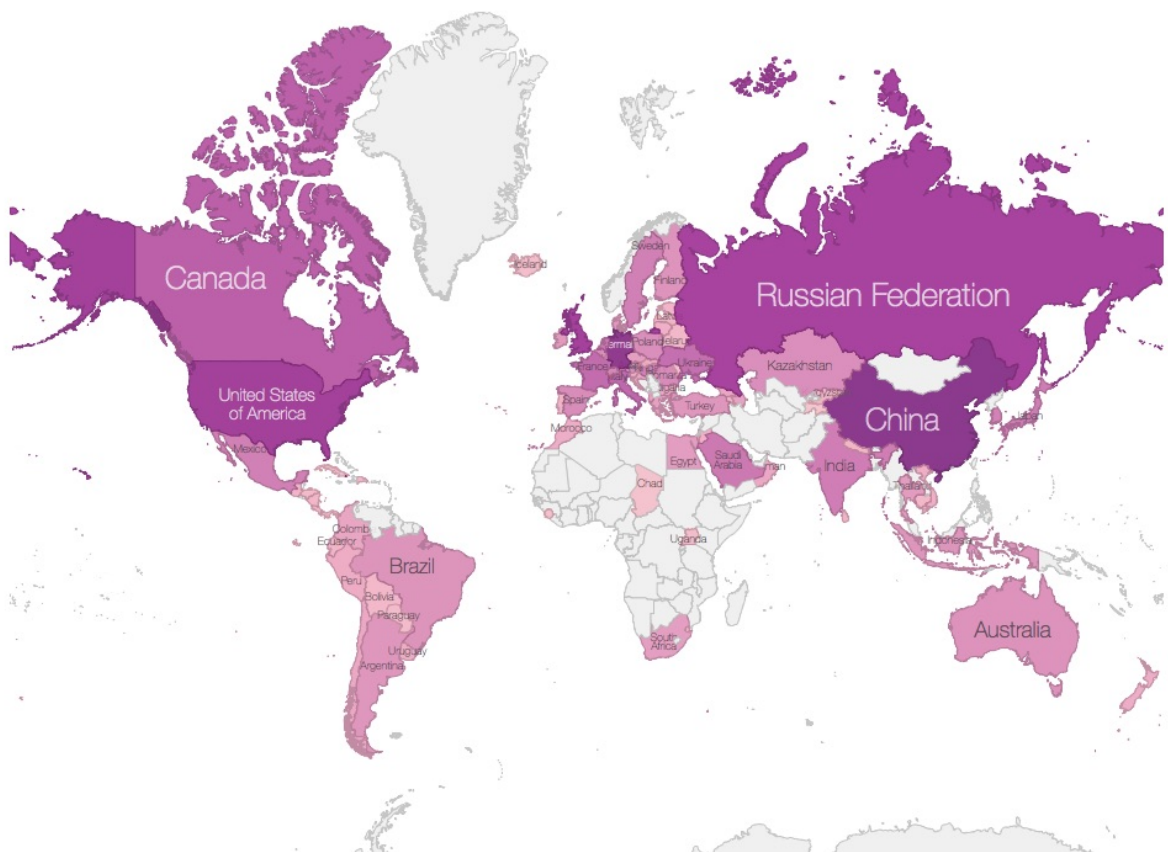
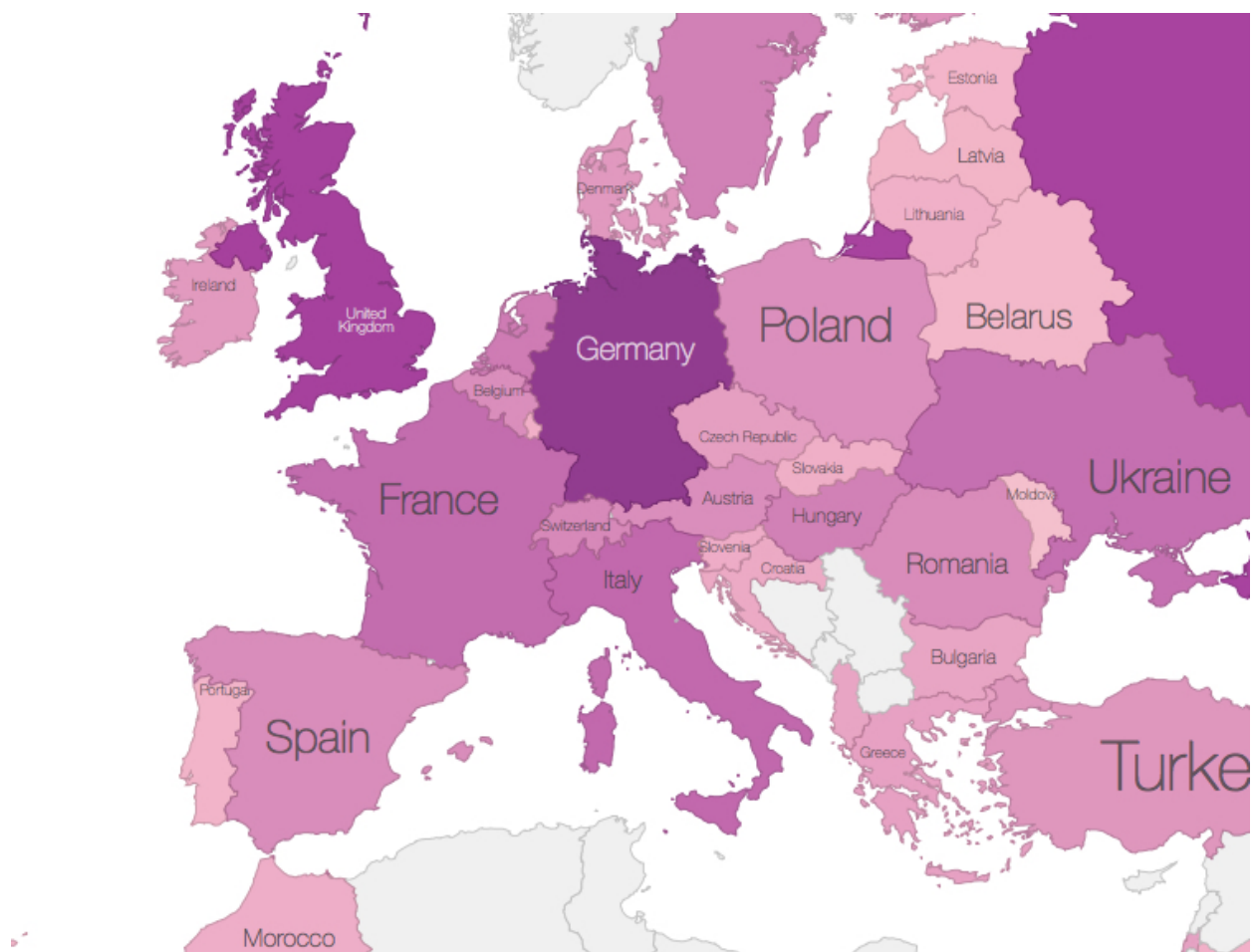Figure 8. Cloropleth of world's outbound tourism, 2014

Figure 9. Cloropleth of Europe's outbound tourism, 2014

References

[1] J. MACKINLAY, "Automating the design of graphical presentations of relational information," *ACM Transactions on Graphics*, vol. 5, no. 2, 1986.

[2] (). 1998–2002 argentine great depression, [Online]. Available: `https://en.wikipedia.org/wiki/1998%E2%80%932002_Argentine_great_depression`.

[3] (). Late 2000s great recession, [Online]. Available: `https://en.wikipedia.org/wiki/Great_Recession`.

[4] (). 1998–2002 argentine great depression, [Online]. Available: `https://en.wikipedia.org/wiki/Early_2000s_recession`.

[5] (). 1998–2002 argentine great depression, [Online]. Available: `https://en.wikipedia.org/wiki/Dot-com_bubble`.