TraveLens: An application that visualizes the World Economic Forum's annual report on Travel and Tourism Competitiveness 2013

Nick Anderson
Lingshuang Chen
Jina Suh
Ruiyi Zhou

Abstract 2
Introduction 2
Previous work 2
Design Process 4
Iteration 1 4
Iteration 2 6
Iteration 3 7
Redesign 9
Discussion 9
Implementation and tools 9
Data sets 10
Visualizations 10
Interaction 12
Color 17
Future work 18
References 19
Acknowledgements 21
Appendix A (TraveLens application) 21
Appendix B (evaluation materials) 21
Appendix C (user research materials) 26

Abstract

TraveLens is an interactive visualization that supports insight about international travel to 140 countries, aimed at the traveler, researcher and investor, across a broad range of economic and demographic indicators describing the global travel and tourism sector. The application visualizes aggregate country scores and their underlying measures from the World Economic Forum Travel and Tourism Competitiveness Report 2013, contextualized by international tourism arrivals data from the World Bank from 1995-2013.

The application presents a dashboard view with a persistent choropleth, scatter plot, and parallel coordinate plot. Users can compare countries across high-level aggregate scores relating to travel and tourism, or focus on a single aggregate to look at its underlying measures within a topic area and investigate clustering or correlation related to those measures. At any point, users can posit relationships between the scores or underlying measures and the overall pattern of tourist arrivals in a country over time. To further expose relationships or clustering, users can compare measures from different categories by choosing a subset and re-plotting them in parallel coordinates that persist in an independent tab, making it possible to develop regional profiles or find the geographic distribution of a category cluster.

Introduction

Our initial project goal was to visualize a combined dataset sourced from the World Bank, World Health Organization, Centers for Disease Control, and other agencies using indicators that might influence travelers' decision-making process. We categorized the datasets by topic--physical safety, health risk, travel cost, and political factors--and attempted to gauge topic influence on potential travelers. However, our initial user feedback was mixed, and visualizing some datasets that measured risk factors tended to encourage assumptions about the economic and cultural position of travelers, rather than explaining or debunking them.

We discarded our original datasets in favor of a highly-structured group of indicators describing travel and tourism compiled by the World Economic Forum, with the goal of allowing users to see these indicators against overall tourist visits. Although it doesn't visualize specific city data or list attractions in ways that might satisfy the typical tourist, we think this design allows users to pursue a range of travel-related comparisons using country dimensions specific enough to support individual itineraries. Users of TraveLens can still view scores for safety and security, but these now appear with reference to economic, policy-making, and survey measures.

Although the WEF report primarily serves an investor and business audience, its underlying data is compiled from World Bank and international agency sources rather than the corporate sector, so we think it can support the goals of individual travelers or researchers as well as investors.

Previous work

Various visualizations influenced either our early design ideas or later design choices; we describe several significant examples here. Influential papers and research appear in References.

World Travel & Tourism Council Annual Economic Impact Report 2014. Visualizations that compare country economic and demographic measures are easy to find on NGO and think-tank websites or included in their annual reports. This example uses an animated 3-D globe for

country selection. Users can compare pairs of countries in the globe dashboard, or view the information in a sortable list and filter by physical or political categories. We were struck by the visual impact of its globe animation, but it is slow and distracting, and allows overlapping of colors to cause distortion of data in the choropleth. However, the visualization offers easy filtering by region, which we've included in future work for our visualization.

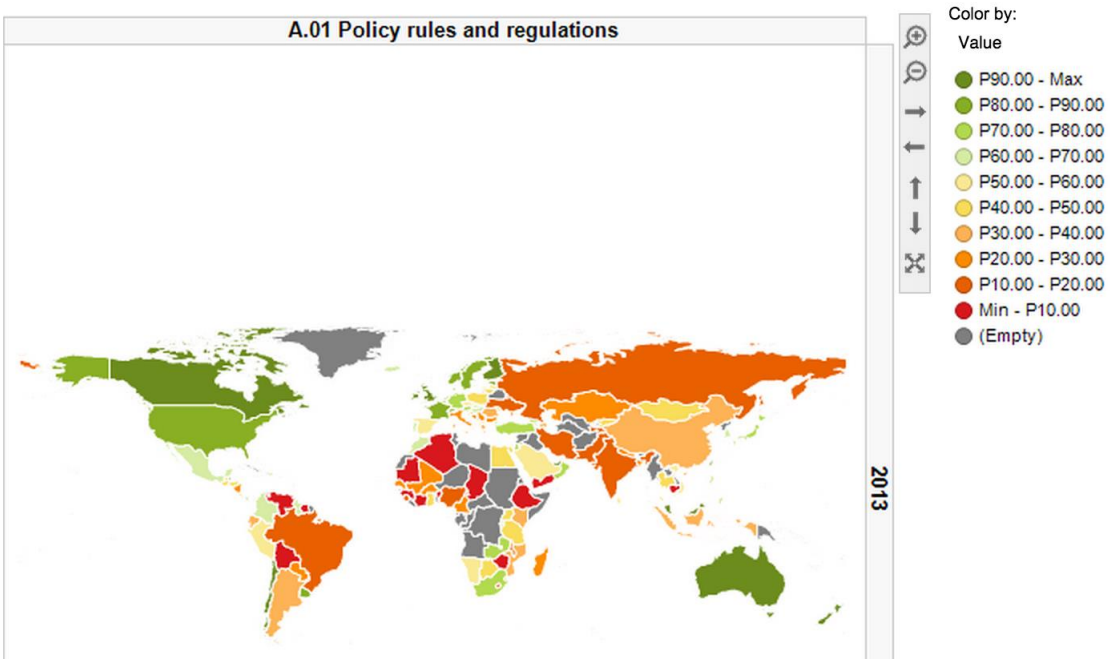

Figure 1: WTTC heatmap

Travel and Tourism Competitiveness Index 2013 Data Analyzer, World Economic Forum. This tool visualizes the report data used for our own project, though we didn't identify this visualization until midway through our design process. Our project differs from the WEF tool in several respects. TraveLens doesn't attempt to interpolate the WEF aggregate scores between 2008, 2009, 2011, and 2013 to provide a time series, as the WEF tool does, because we didn't think the data supported strong time series inferences for our users at that level of abstraction. The WEF tool also encodes quantitative data with color in its choropleth, which imposed a social valuation on the data (red=bad, green=good) that we wanted to avoid. However, the bookmarking, sharing, and export features of the WEF tool are valuable and we include them as future work for our own design.

Nutrient Contents - Parallel Coordinates. This visualization informed some of our decisions about scales, alpha transparency, and viewing clusters in parallel coordinates. We admired, but did not adopt, the curve interpolation it uses for data points.

Parallel Coordinates. Mike Bostock, Protovis. This implementation helped us consider interaction choices for drag selection, especially near axes.

Design Process

Our design underwent several changes in focus and design direction over the course of the project.

Iteration 1

Goals

Early investigations for this project included a range of questions relating to global travel. At first, we imagined an interactive tool that would address the gap between logistical aspects of travel planning (visa requirements, vaccinations, government and NGO alerts) and economic and cultural factors that might inform an individual traveler planning an international trip (political stability, health risks, etc). This vision is represented in our first proposal submission.
Notable goals from this phase included:

- Offering multiple views across multiple indicator categories and decisionmaking criteria, such as health, politics, safety, documentation requirements, natural disasters, travel cost, etc.
- Supporting discovery of unexpected or interesting travel destinations
- Options to plug in other data sets and visualization types

Process

We began investigating data sets in health, traveler safety, weather and climate, and visa/documentation categories. These included Centers for Disease Control health warnings, country visa requirements compiled by other researchers (see References), U.S. State Department per diem allowances and cost of living estimates, and Pew Research survey data concerning attitudes concerning gender equality, religion, and LGBT issues as they pertained to travelers.

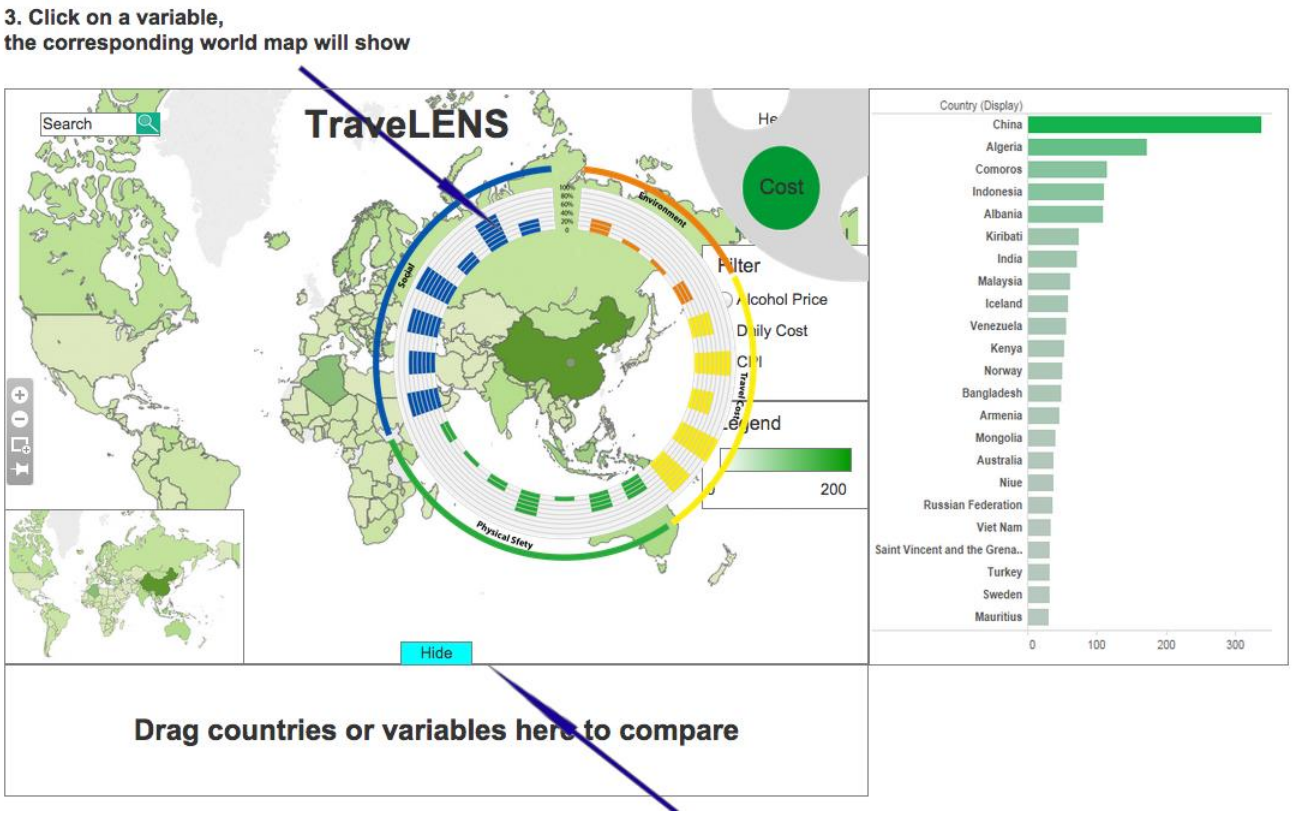

Figure 2: Initial prototype

This was the first prototype we built according to the ideas from ideation (see Fig. 2). It used the choropleth as the main part, supplemented by the bar chart on the right which showed the ranking of the countries worldwide in the selected indicator, the circular bar chart (used the lens metaphor) that showed the scores of all the indicators of the selected country which would popped up when clicking on a specific country, and a comparison zone into which users could drag countries to compare them across all the indicators in bar charts. Aside from choropleth, the bar chart also played a crucial role in this version of prototype.

User Research and Key Findings

User research at this stage was twofold: we used a survey to investigate which indicators describing destination countries were more important to users, and we used an HTML prototype created in Axure to gauge user reaction to the overall design goals of country comparison across aggregate rankings as well as the interaction design.

Users indicated that they were less influenced than we had expected by concerns about cultural differences in choosing a destination country. They indicated that climate and weather information, concerns about crime and physical safety, health problems/disease, and the perceived political stability of their travel destinations were stronger influences in their travel planning. More significantly, users had difficulty comparing different dimensions in the circular bar chart. They were confused by the visual proximity of aggregate dimensions that had no statistical relationship, and they (understandably) wanted to see what measures underlay the aggregates. Some users pointed out that the key indicators (for them) were missing, and wanted others. One reviewer also recommended a simpler design goal: create a visualization that simply helped users discover travel information and make travel decisions quickly, rather than investigate their decision-making.

Next Step

Instead of using a circular bar chart, we looked for another plot type that served our users better, and began considering radar charts and parallel coordinates.

Iteration 2

Goals

Based on initial research, we shifted our design goals somewhat. We retained the goal of visualizing travel requirements and country overview information, but wanted to emphasize travel risk factors. We thought this might support new insights about travel decision-making, since risk perception had emerged as a theme during the first round of usability assessment. Feedback from our mid-term presentation also indicated that users might want to see more detailed information at city level.

Notable goals from this phase included:

- Finding new data sets to support questions about risk as a factor in travel
- Finding better way of visualization than bar chart
- Offering city-level information

Process

We revised our prototype by removing some mocked-up interactions for drag-and-drop country comparison and replacing the circular bar chart with parallel coordinates in a separate dashboard. We also experimented with radar charts to represent subcategory measures that made up the fourteen aggregates. (Figures 3 and 4)

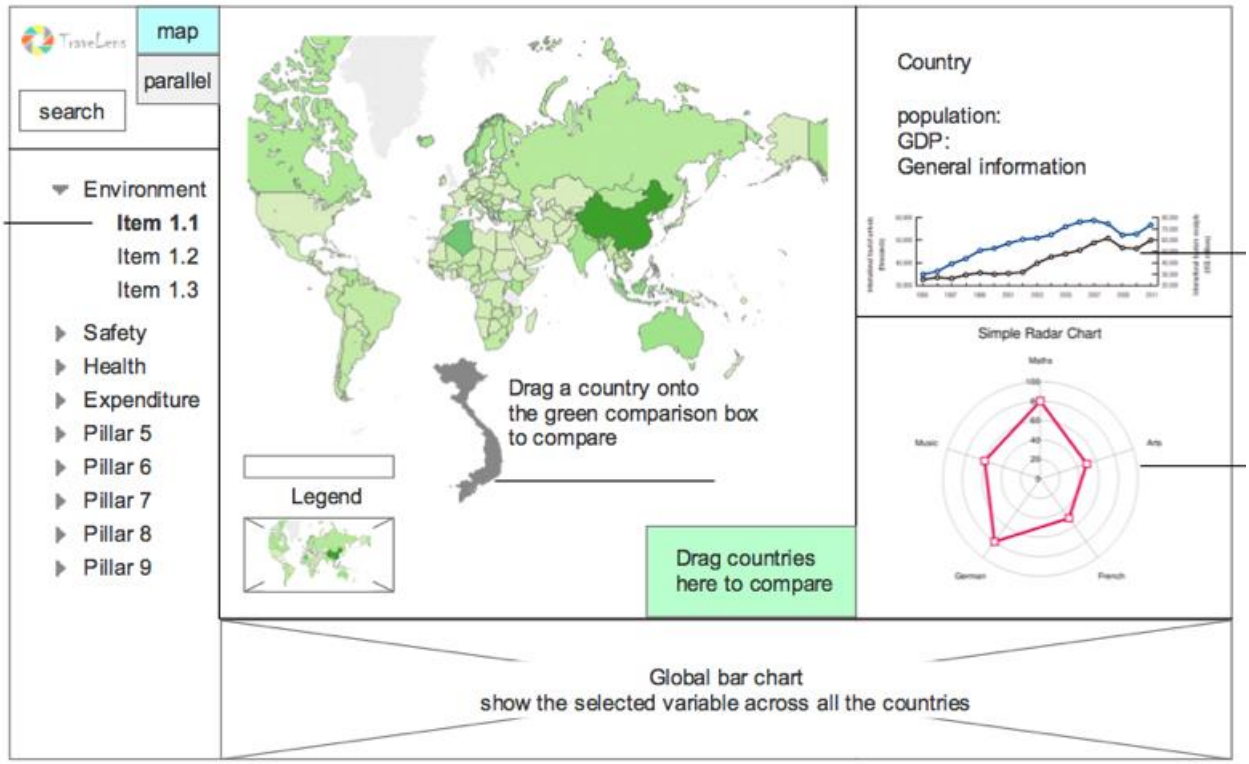

Figure 3: Second prototype

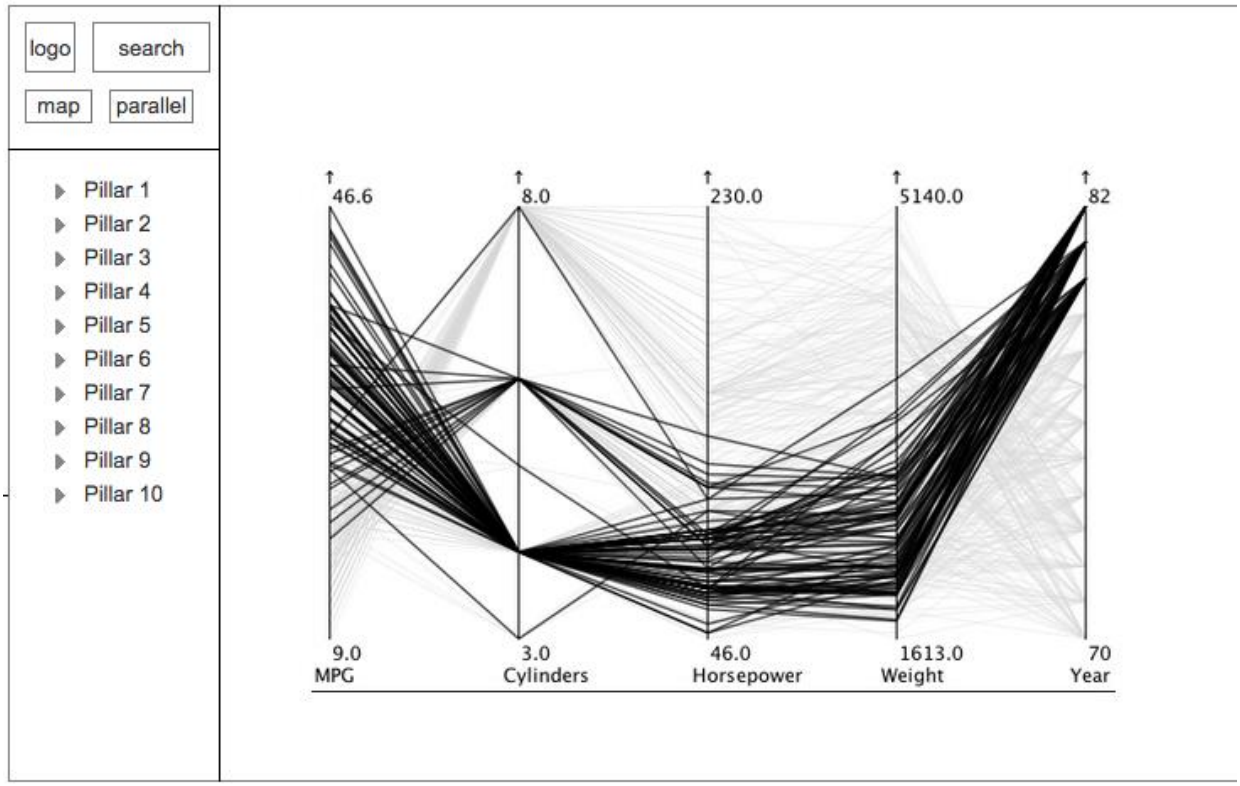

Figure 4: Second prototype, parallel coordinates

User Research and Key Findings

We used the revised prototype for two semi-structured interviews. Our users were engaged by the overall project idea, but pointed out that their travel decisions were highly individual and situational. They valued city and regional information, and they wanted indicators that they perceived as predictive, whereas the data categories we visualized were abstract and descriptive. Our users remained interested in exploring a wide range of indicators about their travel destinations, but they didn't confirm that the data we visualized influenced their actual travel planning. They wanted to see more general country information such as language, currency, and time zone, and they needed more supporting context before understanding how to use the parallel coordinates.

Next Steps

Based on these interviews and our own evaluation, we decided that the inferences we tried to draw about risk factors and travel decisions weren't making sense in the visualization.

Iteration 3

Goals

We decided to narrow our project scope further by providing an open-ended set of comparisons across general travel and tourism indicators, using a mix of aggregated rankings, quantitative measures, and survey data, and leaving specific questions up to our users.

Process

For our third design phase, we committed to visualizing the WEF Travel and Tourism Competitiveness Report; we also wanted to show the report data in context of actual tourist arrivals in the represented countries, which was not part of the WEF report. We evaluated several different sources of tourist arrivals data. For example, we considered using U.S. Travel and Tourism statistics to see specifically where U.S. residents travel, and we looked at inbound and outbound flight data between U.S. airports and international airports. Because some of these datasets included non-tourist data points that we couldn't filter out, we used World Bank measures that were tourist-specific but which combined arrivals by air with overland, marine, and border arrivals. We also added general country information (currency, language, time zone, etc.) to the visualization.

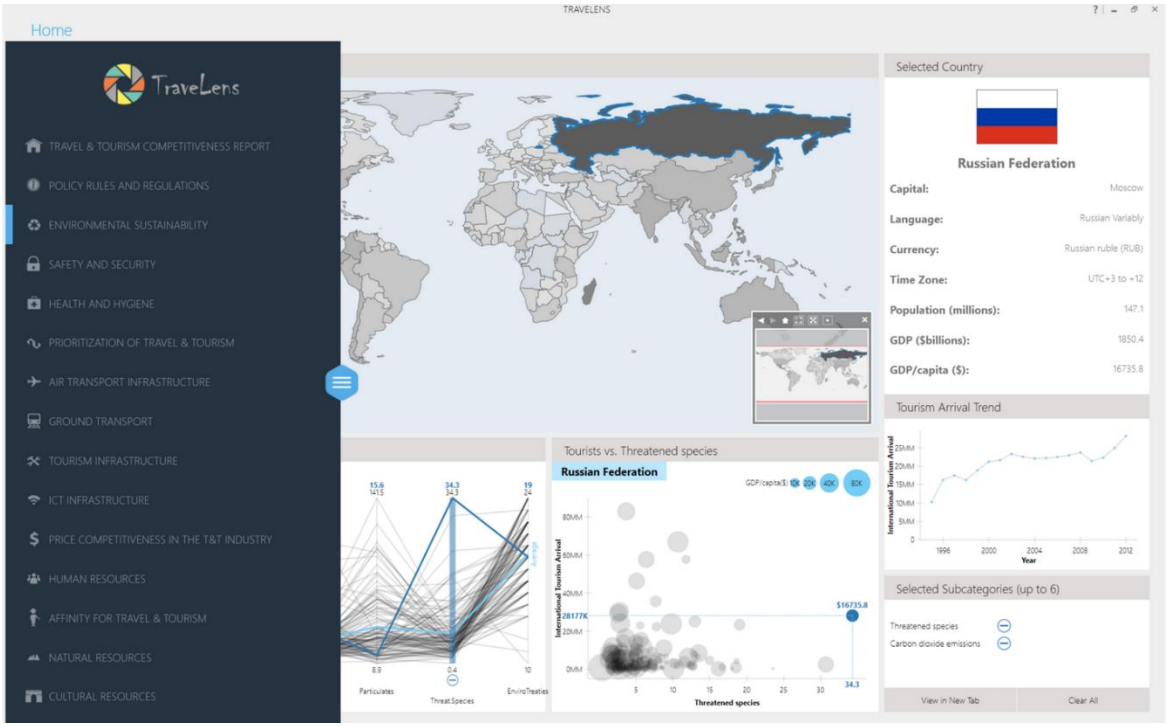

Figure 5: Current application

In the third version of design (Figure 4), we moved the parallel coordinates from the second tab to the main tab, in order to offer more insights using brushing and linking, and we added the scatterplot using the selected category as the x axis, the international tourism arrival as the y axis, and the area encoded the GDP/capita. On the bottom right, we added the comparison zone, the categories users chose would show up here and users could get a new tab comparing these categories.

User research and key findings

As soon as we finished building the visualization, we conducted two rounds of task-based interviews with users and did some heuristic usability assessment. We confirmed that users understood the main idea of the tool, could explore the data categories, and could perform at least some discovery tasks.

- We found that users needed help recognizing subordination in the data (distinguishing topic categories and their scores from the subcategory measures).

- We needed to clarify units of measure on all parallel coordinate axes, and provide detailed descriptions of subcategory measures.
- Users didn't always recognize that the the heatmap reflected aggregate scores for the categories.
- Users had various reactions to our implementation of multiple selection; the most important issues were difficulty distinguishing between lines and making individual selections (in parallel coordinates) and value labeling for multiples.
- Users responded to the "profile recognition" aspect of using parallel coordinates, but had the typical difficulty examining relationships between axes that weren't adjacent.
- Users looked for an interactive legend for data display and selection, and expected to be able to interact with the choropleth legend as well.
- Users were interested in a quick tutorial function.

Redesign

We focused on several primary design changes: improving interactions around multiple selection, consistent labeling and values display, better subcategory comparison, and improving user understanding of the measure/aggregate relationship of the data categories.
Improving multiple selection was crucial and tricky. At the time of our presentation, the application used one color to indicate focus and another to indicate selection, with labeling changing on mouseover of each record in the parallel coordinates and scatterplot, or on mouse entry of an outline on the choropleth. For single selections, this let the user compare the selected record with any record in focus, but when users drag-selected multiple coordinates, they could only identify their selections by returning focus to that record.

We added control-selection and a palette of colors to differentiate multiple selections. To improve visual comparison of clustered records in the coordinates, we dimmed the unselected records when the user selected more than one, thinking that it was more important to see selections clearly than to see any surrounding clustering. Finally, we added an interactive countries legend coordinated with the other plots, and persisted paged country details for selected countries. See Discussion for a fuller description of this interaction.

To improve users' ability to examine relationships in the parallel coordinates, we added functionality to let users select axes (in the user's preferred order) and generate new dashboard tabs that plotted their selections. See Discussion for more on this design choice.

To improve user orientation in the data categories, we added breadcrumb labeling to the choropleth title and a "category/subcategory" naming convention to the parallel coordinate window. Axis labels now stayed in view all the time, not only on mouseover, and appeared below the choropleth legend as well to cue the reader about aggregate scores.

Finally, we wrote and implemented a brief in-app tutorial that stepped through basic info discovery and selection.

Discussion

Implementation and tools

We went through several implementation options during our investigation into visualization tools: Tableau, d3, and WPF (Windows Presentation Framework written with .NET and XAML). We used Tableau extensively for initial exploration of potential datasets. Our general consensus was that Tableau is very useful in data exploration, but it could not create some of the

visualizations we wanted (especially the circular bar chart), it lacked the UI toolkit to create an overall application experience, and it slowed significantly when joining multiple datasets with large data points.
d3, on the other hand, was highly optimized for data visualization. During initial investigation, we encountered a few open-source libraries that supported choropleth visualization in d3 with zooming capabilities and geoJSON data format, and we visualized one of our many datasets in a working d3 prototype. Our main concerns with these controls were with performance as well as interactions. Because no member of our group was adept with javascript, we concluded that debugging d3 controls to work with all of our data and add custom interactions (brush-link, menus, focus+context, etc.) would be highly risky, especially because we had uncertainties in our dataset.
We chose to develop our project in WPF for several reasons: we were fairly proficient with graphics primitives in WPF; WPF supported interaction and data binding well; and we understood the language well enough to be able to optimize for performance. Choosing .NET as a programming language was extremely helpful with data cleanup tasks because .NET natively supported SQL operations. In most of our iterations with many datasets, we used .NET and SQL to parse the data, dump the data into tables, perform joins and aggregates, and export the dataset to be visualized in Tableau, d3, or WPF.

Data sets

TraveLens includes 14840 data points.

- WEF data set--14 aggregate dimensions with 79 underlying measures for 140 countries.
- World Bank international tourism arrivals from 1995-2012 for 140 countries.
- WikiTravel country details for each country (9 nominal dimensions displayed in tabular form).
- To create effective tables, we mapped country names to ISO 2-letter and ISO 3-letter country codes using Wikipedia's pages for ISO standards.
- Our choropleth uses geoJSON data to draw country outlines at 1:50 scale.

The World Economic Forum's annual report data set subsumed many of the data sets we had initially investigated from WHO, UN, and NGOs. Other advantages:

- This data set supports scenarios for individual travelers as well as investors and the travel industry.
- WEF did prior work to adjust weighting and clean the data set (unlike our earlier data sets).
- The set is highly organized and subordinated; this structure helps us offer broad overview while allowing the user to investigate the data underlying the aggregate scores.

Visualizations

Choropleth

Our design goal is to allow users to compare variables across countries. While there are other fine choices for visualizing data comparison (e.g. bar chart), our users indicated that geospatial context is critical for locating groupings across countries and maps make the most sense to travelers. We chose to visualize the selected variable in the choropleth because it easily shows variabilities regionally as well as globally.
Our choropleth visualization includes two additional components that are worth noting: a minimap and an interactive legend. The minimap provides focus and context so that the user can zoom into a region of interest and see the context at the same time. Because a heat map does not allow for the user to extract the exact value represented by the color, we added an interactive legend which reveals the value for the selected or hovered country.

Parallel coordinates

Our design relies heavily on parallel coordinates for user comparison because this plot type supported our particular set of design goals, user tasks, and data set characteristics so well.

Most importantly, we wanted to support comparison across many dissimilar variables, a common case for parallel coordinates. Many of the measures within a topic (Internet and Communications Infrastructure, for example) are aggregates that might include survey scores, a percentage of budget, a population count, or a percentage-of-benchmark score.

Because our basic unit of comparison is the country, we also speculated that countries might show a recognizable visual profile across axes, and that this profile might be easier to associate with a country than with a more arbitrary sort of record. Per Stephen Few, "...most often when analyzing multivariate data, we look for multivariate profiles that correspond to a particular condition." (*Now You See It*, 289)

To address common design problems with parallel coordinates, we specified partial opacity for values to reduce occlusion, drew values in color only for selection, drew a line to indicate average across variables, and provided detailed labeling for axes that was coordinated with our other plot types. We also supported user selection and re-ordering of axes to support better insight into variable relationships and clustering.

We wanted to visualize the subcategory measures in the same way that we visualized the entire set of 14 aggregate scores. Our only other option here would have been scatterplots as small multiples, but we opted not to do this. Multiples would have been so small (with 140 points each) that occlusion would have been a real problem, and would not have made the cognitive task of comparison easier (in our view).

Scatter plot

Given a popular tourist destination, users want to find out what factors make it popular; given tourism factors, users want to find out if they have an impact on tourism. We needed to provide a way to tie a variable of interest with the actual tourism outcome. Scatter plots were the best choice for visualizing potential correlation between international tourism arrival and a variable of interest. They reveal potential outliers immediately, and they support the scatterplot with a more familiar plot type for most users.

Because our dashboard uses relatively few plot types, we made a few design trade-offs with the scatterplot. Some variables don't readily compare with tourist arrivals (rate of treaty ratifications, business-friendliness measures). We also visualize GDP/capita in the scatterplot, regardless of the variable in the X-axis. We did this chiefly because overall economic prosperity in a given country was the indicator that we had to "rule out" most often when we looked for correlations in the data set. We thought persisting users' view of GDP/capita would make it easier for them to double-check correlations that seem to appear in the heatmap or parallel coordinate, and it would clarify some variables that are calculated as percentage of budget.

Country detail section

Based on user feedback from our second iteration, we incorporated the country detail area to extend the tourism data. In the top section, we list common guidebook facts about a country (e.g. language, flag, currency, etc.). In the bottom section, we provided a time series of tourism arrivals for the given country from 1995 to 2012 so that the user can assess the currently selected variable from the perspective of the historical tourism trend. For variables specific to tourist visits (tourist infrastructure, police effectiveness), users might want to know whether they're planning a trip on a trend upswing or down slope.

Interaction

We learned early on in our user research and evaluation that the travel-related decision making process can be highly individualized. It is often difficult to anticipate all use cases especially when dealing with data analysis, and our dataset was broad in the range of topics it covers. Rather than focusing heavily on use cases based on a few insights we have discovered in our dataset, we concentrated our efforts on providing rich interactions for our users to help our users make their own discoveries and insights based on key tasks such as comparison of variables and selection of countries.

Another key motivation behind extensive interactivity is the user confusion behind parallel coordinates. Parallel coordinates are useful for illustrating the "big picture" and drawing out the patterns and exceptions across many variables. However, because parallel coordinates often involve many lines and occlusion of these lines, it is hard to distinguish each data point or separate data points that are of interest to the users. Interaction techniques such as brushing, filtering, and selection can help the users navigate parallel coordinates.

To critique our interactive visualization, we considered Heer and Schneiderman's taxonomy of interactive dynamics for visual analysis. Of the twelve tasks defined in the taxonomy, we discuss in detail four important tasks that we believe are key ingredients of TraveLens: Filter, Selection, Coordinate, Organize.

Filter

Our dataset contains 94 variables with hierarchical structure within those variables, and we support inspecting each of these variables by the use of filters. In order to illustrate this hierarchical structure, the 14 top level categories are exposed in the drawer. When the user selects one of the top level categories, we expose its subcategories in the parallel coordinates. Users can select the top level category from the drawer and update all of the data visualizations to show that category's data. Once a top level category is selected, users can select each of the subcategories by clicking on the axes of the parallel coordinates to show the subcategory data. Therefore, the drawer acts as a filter for the high level concepts (e.g. Natural Resources), and the axes of the parallel coordinates act as filters for the detailed level concepts (e.g. Number of Endangered Species).

We placed the top level categorical filters in the drawer rather than exposing a list of checkboxes within the visualization because (a) they are global filters that update all of the visualizations at once, and (b) Home is used for exploratory data analysis where the users are looking at a specific travel-related category at a time.

We placed the subcategory level filters in the parallel coordinates rather than in the drawer so that, within a category, the users can switch between subcategories quickly while seeing the context of how that variable relates to other variables. We experimented with placing the subcategory filter as a list of checkboxes within the visualization, but doing so introduced user confusion and inconsistencies between visualizations because the parallel coordinates cannot be used to visualize one variable at a time while choropleth and scatterplot (with y-axis fixed to tourist arrival data) can visualize only one variable at a time.

Selection

Our dataset contains 140 countries. There are two user tasks involving countries when we consider the travel decision-making process. One is to look at all of the countries at once to discover patterns and to search for countries that meet a certain set of user criteria; another is

to look at only a selected subset of countries and to compare those countries across variables. We support these two tasks by the use of selection.

In our choropleth, scatter plot, and parallel coordinates, the user can directly select countries by clicking a single country or by performing click-drag selection (e.g. “lasso”) or control+click to select multiple countries. We tried to mimic the tried-and-true interaction behavior of typical OS desktop as much as possible. In the choropleth, the usual task involves selecting countries by region of interest or a list of specific countries in mind, so the selection is motivated by the geographical location of countries. When the user is in the process of selecting countries inside the choropleth, we don’t dim the other unselected countries because all countries have equal importance. In the scatter plot and the parallel coordinates, the user normally selects countries based on where the countries fall along the given variable, so the selection is motivated by pattern recognition. When the user has selected countries, we purposefully dim the other unselected countries in order to enhance the pattern recognition experience. This ability to see patterns clearly is one of the key usability feedbacks we received during our usability evaluation and incorporated into the final visualization.

Coordinate

Our choropleth serves as a heat map of the selected variable as well as a geographical navigation aid. Our parallel coordinates plot serves as a comparison and correlation tool for multiple variables simultaneously. Our scatter plot serves as a correlation tool that contextualizes the selected variable against the baseline tourist arrivals data. Because these three visualizations play a crucial role in telling a story about tourism from several key perspectives, coordinating this multi-view data analysis was one of our highest priorities, which we support by the use of brushing and linking. We also support coordination through the use of colors which we will discuss in later section.

As discussed in the previous section, the user might select countries based on their geographical location and see the correlation between multiple variable in the parallel coordinates and scatter plot. The user might select interesting patterns in the parallel coordinates and see where those lines (or countries) are located geographically and how they relate to the baseline tourism indicator. The user might select unusual outliers in the scatter plot and see where those countries are located, how they related to other variables, and if they are outliers across all subcategory level variables.

The use of brushing and linking in coordinating multiple views is such a powerful technique in our interactive visualization that reveals potentially hidden insights. Fig. 6 illustrates a scenario where the user was interested in Brazil as a candidate travel destination from the choropleth. The user can immediately see where Brazil stands across multiple subcategories in the parallel coordinates. The user can also see how the selected variable correlates to the tourist arrival data for Brazil in the scatter plot. Fig. 7 illustrates a scenario where the user examines the outlier for the selected category in the scatter plot, and the coordination reveals the country geographically in the choropleth and shows how that outlier performs across other variables in the parallel coordinate. Fig. 8 illustrates a scenario where the user examines countries that intersect below the top portion of an axis in the parallel coordinates. The choropleth reveals where those countries are clustered geographically, and the scatter plot shows how those countries for the selected variable correlate to the tourist arrivals.

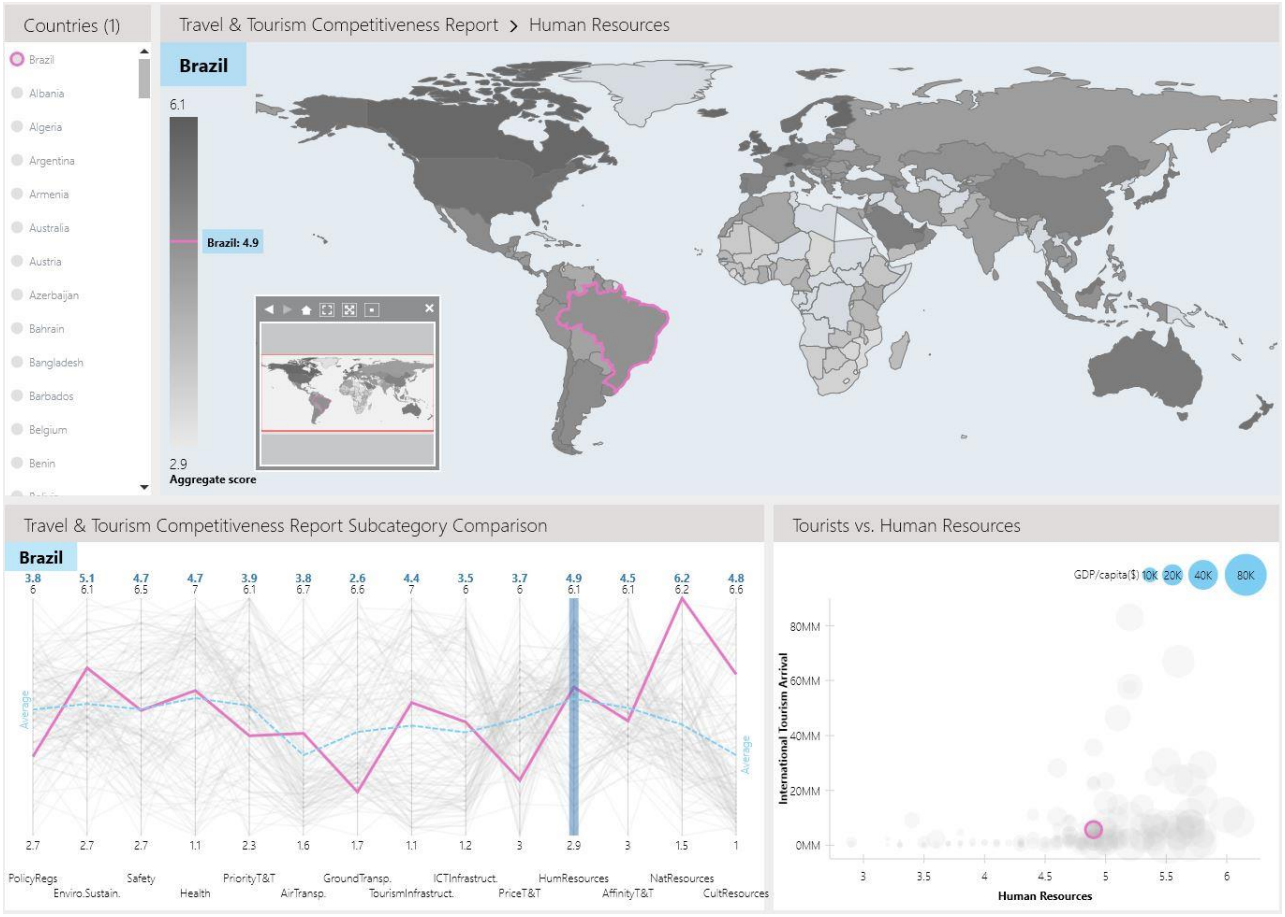

Figure 6: Country selection

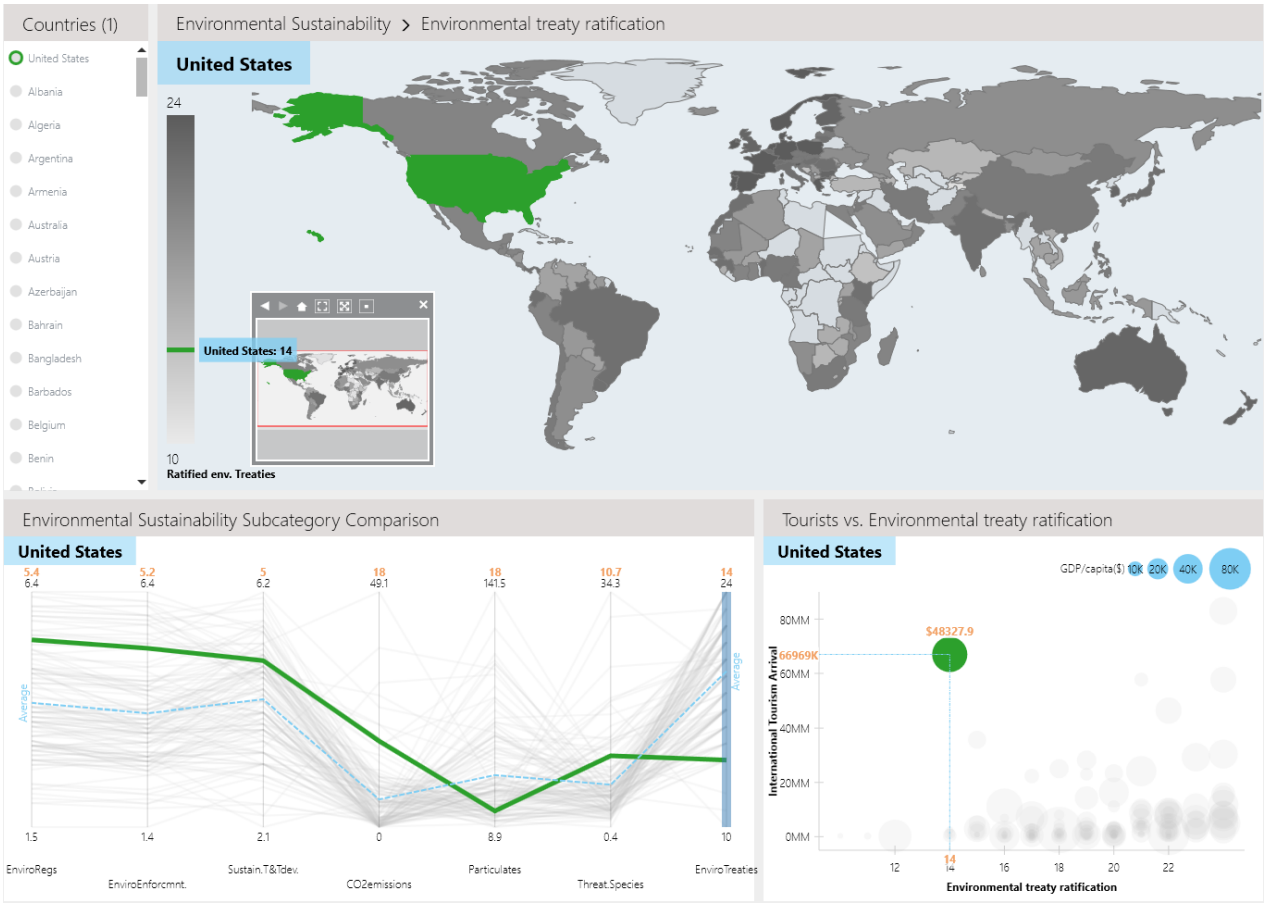

Figure 7: Outlier comparison

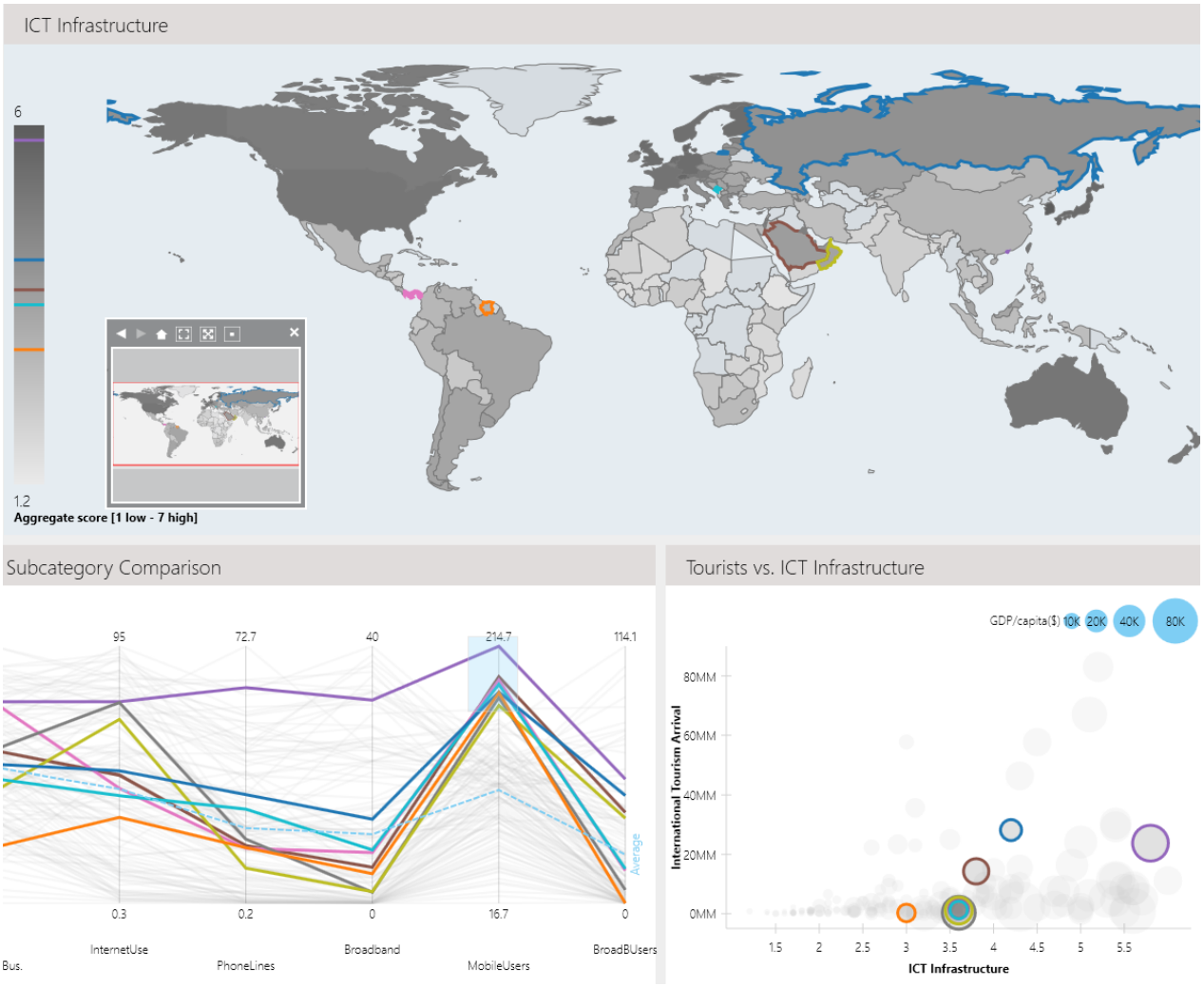

Figure 8: Cluster selection

When the user has selected countries, we disable hover-over on unselected countries in the scatter plot and the parallel coordinate to simplify interaction. We concluded, based on usability evaluation, that the users interact with the bottom two visualizations for two main tasks: (a) selecting countries from the choropleth and then examining patterns and correlations in other visualizations, and (b) selecting patterns from the parallel coordinate or the scatter plot and then seeing which countries fit those patterns. In either of these scenarios when the users are closely examining a selected subset of the data points, having hover-over interaction across our visualizations introduced too much visual clutter in the user's peripheral vision and interfered with the detailed inspection of the data points that the users cared about. Parallel coordinate, by itself, was visually and cognitively intensive to the users, so simplifying this interaction became our highest priority, and, therefore, we removed unnecessary hover-over highlights when the users were inspecting a selected subset of countries.

We made further improvements to add consistency in coordination across multiple views. In our previous prototype, we marked the selected countries with a border in the choropleth, a line in the parallel coordinate, and a filled circle in the scatter plot. In our final visualization, we marked all selected data points with borders to avoid this visual inconsistency, to reduce the amount of occlusion in the scatter plot for overlapping data points, and to reveal the actual value encoded

by the grayscale in the choropleth. We chose to fill the hovered area with the selection or hover color to emphasize the selection or hover-over and to make it easier to find the hovered data point across multiple visualizations.

Organize

We had initially only allowed the users to compare variables within a selected category. For example, if the user selected Safety and Security, we only allowed them to compare subcategories within Safety and Security. We immediately experienced that most insights involved comparison between subcategories across multiple categories. It was cognitively demanding to remember analysis results from the previous categories, and our users confirmed these observations.

We experimented with different ways to allow users to select multiple subcategories (e.g. list of checkboxes for 94 variables), but browsing through such long list of variables was tedious and demanded categorization which we previously had. We observed that the users first *explore* different categories and its subcategories, and only after learning about the subcategories that matter to them, they needed a way to *analyze* those subcategories side-by-side. We saw that the users performed their tasks in this two-step process, so we support organizing the interactive visualization into tabs. During the exploration phase, the user remains in the home tab where the categorical hierarchy is fixed. Then, in the analysis phase, the user can select several subcategories of interest and compare them in separate tabs.

We support selection minimum of two subcategories because we need at least two variables for comparison. We support selection maximum of six subcategories mainly for the interest of time, and viewing six variables side-by-side was the most optimal visual configuration for the parallel coordinates. Several additional interactions can be integrated to enhance the analysis experience, and we will discuss these in the future works section.

Color

Color was an important topic of discussion because our tools did not come with a default set of colors for data visualization, and dealing with data points from 140 countries across 94 variables meant that we had to pay an extremely close attention to how we highlight potential insights. For color selection, we wanted to create a mood that is calm and easy. We chose gray as the theme color, using a few bright colors to indicate labels, focus, and selection. These hues are borrowed from the palette in Tableau. The gray works well to provide a context while the bright colors provide the highlights and details. We experimented with color (green) in the choropleth, but found that it didn't improve users' recognition of the heatmap and distracted them from the parallel coordinates.

We originally used a single color (blue) to indicate selected countries, but our users wanted the ability to identify countries without hovering over individual data points. While color is a natural choice for encoding nominal variables, it wasn't an effective encoding for distinguishing between 140 different countries. Given the situation that too many distinct colors would make the identification less effective, and that selecting and making comparison among more than 10 countries is a less-common case, we provide a total of 10 distinct colors, extracted from Tableau 10, to encode the selected countries. The assignment of colors is decided by the sequence of choice instead of predetermining colors for each country to minimize repetition of colors (see Fig. 9).

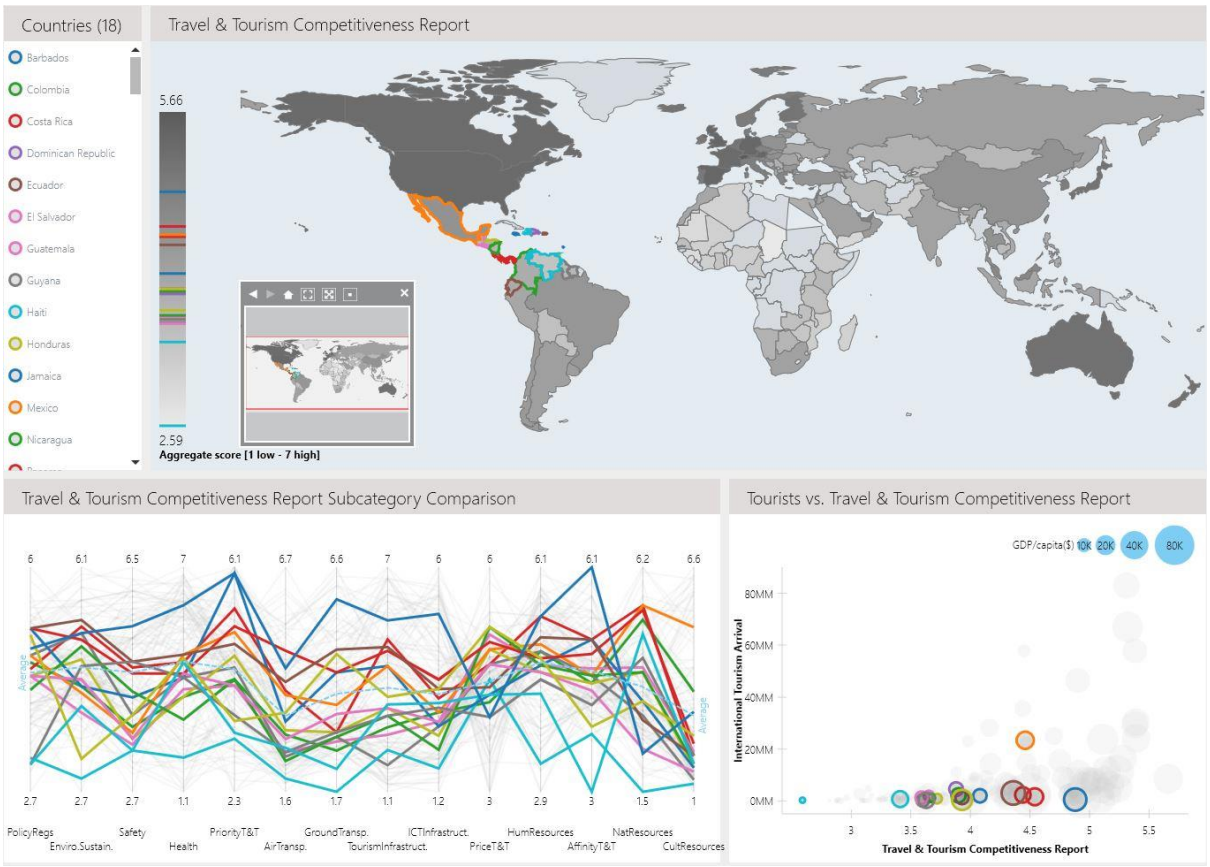

Figure 9: Selection color ordering

Future work

As with most projects, our aspirations were limited by time and feasibility. Following are some aspects of this visualization that we would add or extend in future.

Additional time series

The WEF report is not issued annually (and has changed its data dimensions somewhat between issues). We would like to visualize additional data sets that extend several of the more interesting WEF report categories--communications infrastructure, social factors, and environment indicators--over time.

Support for "extract" task

Per Schneiderman, we would like to support user export and sharing of selections and comparisons, especially when users have generated additional tabs. Because we didn't implement TraveLens as a web application, users can't simply send or bookmark an HTTP address with query parameters in order to share their work. However, the app could export a configuration file for a specific visualization, or export SVG viewable in other applications or web pages.

Tabular filtering for parallel coordinates

We liked the tabular records view in the Nutrient Contents visualization, and would add a similar view for country filtering.

Functionality improvements and bug fixes

- Higher-fidelity choropleth (our geoJSON file has some misdrawn outlines that we haven't been able to fix or replace).
- Support for re-ordering, adding, or removing axes in parallel coordinates.
- User selection from 94 variables for scatter plot X- and Y- axes.
- Optional switching of scatter plot Y-axis scale from linear to log.
- Link to guidebook-style travel website for each country.

References

1. Battermann, Joerg. GeoJSON.NET (code project). Retrieved from https://github.com/jbattermann/GeoJSON.Net/blob/master/src/GeoJSON.Net/Properties/AssemblyInfo.cs
2. Bostock, Mike. "Parallel Coordinates." Retrieved from http://bl.ocks.org/mbostock/1341021
3. Few, Stephen. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Oakland, Calif: Analytics Press, 2009.
4. Foresti, Stefano, James Agutter, Yarden Livnat, Shaun Moon, and Robert Erbacher. "Visualization for Cybersecurity - Visual Correlation of Network Alerts - the Visalert Visual Correlation Tool Facilitates Situational Awareness in Complex Network Environments by Providing a Holistic View of Network Security to Help Detect Malicious Activities." *Ieee Computer Graphics and Applications*. 26.2 (2006): 48. Print.
5. Heer, Jeffrey, and Ben Schneiderman. "Interactive Dynamics for Visual Analysis." *ACMQueue* 10:2 (2012). Retrieved from http://queue.acm.org/detail.cfm?id=2146416.
6. HtmlAgilityPack. http://htmlagilitypack.codeplex.com/
7. Hsu, A., J. Emerson, M. Levy, A. de Sherbinin, L. Johnson, O. Malik, J. Schwartz, and M. International Monetary Fund, International Financial Statistics and data files (2014), Retrieved from http://data.worldbank.org/indicator/FP.CPI.TOTL
8. Jaiteh. (2014). The 2014 Environmental Performance Index. New Haven, CT: Yale Center for Environmental Law and Policy. Available: http://www.epi.yale.edu.
9. Json.NET http://james.newtonking.com/json
10. Kaly, U.L., C.R. Pratt, and J. Mitchell. 2004. The Environmental Vulnerability Index (EVI) 2004. *SOPAC Technical Report* 384, 323 pp.. http://www.sopac.org/sopac/evi/Files/EVI2004TechnicaReport.pdf.
11. Kosara, Robert. "Parallel Coordinates." Retrieved from https://eagereyes.org/techniques/parallel-coordinates
12. Kozak, M, J C. Crotts, and R Law. "The Impact of the Perception of Risk on International travelers." *International Journal of Tourism Research*. 9.4 (2007): 233-242. Print.
13. Lepp, A. "Tourist Roles, Perceived Risk and International Tourism." *Annals of Tourism Research*. 30.3 (2003): 606-624. Print.
14. Mandelkow, Torsten. "Modern UI Charts Demo." (Code project) Retrieved from https://modernuicharts.codeplex.com/
15. Meinshausen, Paul, and Aarti Saxena. *The Index of Tourist Inequality*. Retrieved from http://housesofstones.com/touristinequality.html
16. Neumayer, Eric. "The Impact of Political Violence on Tourism: Dynamic Cross-National Estimation." *The Journal of Conflict Resolution*. 48.2 (2004): 259-281. Print.
17. "Nutrient Contents - Parallel Coordinates." Retrieved from http://exposedata.com/parallel/
18. Pew Research Global Attitudes Project. (2014). *Spring 2013 survey* [data, questionnaire, and topline results]. Retrieved from http://www.pewglobal.org/2013/06/04/the-global-divide-on-homosexuality/.

19. Pew Research Global Attitudes Project. (2014). *Spring 2010 survey* [data, questionnaire, and topline results]. Retrieved from http://www.pewglobal.org/2010/05/08/spring-2010-survey-data/.
20. Quintal, Vanessa A, Julie A. Lee, and Geoffrey N. Soutar. "Tourists' Information Search: the Differential Impact of Risk and Uncertainty Avoidance." *International Journal of Tourism Research*. 12.4 (2010): 321-333. Print.
21. Reisinger, Yvette, and Felix Mavondo. 2005. "Travel Anxiety and Intentions to Travel Internationally: Implications of Travel Risk Perception". *Journal of Travel Research.* 43, no. 3: 212-225.
22. Schneiderman, Ben. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In Bederson, Benjamin B., and Ben Shneiderman, *The Craft of Information Visualization: Readings and Reflections*. San Francisco, Calif: Morgan Kaufmann, 2003.
Sharifpour, M, G Walters, B W. Ritchie, and C Winter. "Investigating the Role of Prior Knowledge in Tourist Decision Making: a Structural Equation Model of Risk Perceptions and Information Search." *Journal of Travel Research*. 53.3 (2014): 307-322. Print.
23. Sönmez, S. F., and A. R. Graefe (1998a). "Determining Future Travel Behavior from Past Travel Experience and Perceptions of Risk and Safety." *Journal of Travel Research*, 37: 171-7.
24. South Pacific Applied Geoscience Commission - SOPAC, and United Nations Environmental Program - UNEP. 2004. Compendium of Environmental Sustainability Indicator Collections: 2004 Environmental Vulnerability Index (EVI). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H45H7D60. Accessed 10/16/2014.
25. Sundstrom, Johan. World.geo.json (code project). Retrieved from https://github.com/johan/world.geo.json
26. United States Department of State. (2014). *Foreign Per Diem Rates by Location*. October 2014 [data]. Retrieved from http://aoprals.state.gov/content.asp?content_id=233
27. WikiTravel, "Quick Facts." Retrieved from 140 country pages, http://wikitravel.org/en/
28. World Bank. "International tourism, number of arrivals". 13 Oct 2014. Retrieved from https://datasets.proquest.com/data/set/15zv/international-tourism-number-of-arrivals#!ds=15zv!i0h&display=line
29. World Economic Forum. "Travel and Tourism Competitiveness Report 2013." Jennifer Blanke and Thea Chiesa, eds. Retrieved from http://www.weforum.org/reports/travel-tourism-competitiveness-report-2013#
30. World Health Organization. Global Health Observatory Data Repository. "Global Status Report on Road Safety 2013." Retrieved from http://www.who.int/gho/road_safety/en/.
31. "World Travel & Tourism Council Annual Economic Impact report 2014." World Travel & Tourism Council. Retrieved from http://wttc-infographic.org/
32. Yavas, Ugur. "Correlates of Vacation Travel." *Journal of Professional Services Marketing*. 5.2 (1990): 3-18. Print.

Acknowledgements

Saleema Amershi
Suzanne Arlt
Ailey Armstrong
Namcy Chan
Jeremy Im
Patrice Simard

Appendix A (TraveLens application)

TraveLens source code (Github project)

Appendix B (evaluation materials)

Evaluation checklist
We solved all the problems that have high severity levels (1 and 2) and some of the problems that have relatively lower severity levels.

Severity 1: An issue that blocks a substantial number of users from using the application. E.g., Users are unable to navigate the app. Nearly impossible task during the initial experience.
Severity 2: An issue that blocks a substantial number of users from making use of an important feature or accomplishing an important task. E.g., Unclear solution. Undiscoverable feature.
Severity 3: An issue that hinders some users from accomplishing a task or goal. E.g., Inconsistent or misleading text. Users don't understand a non--essential aspect of the application or feature. Delays.
Severity 4: An issue that causes some users to become mildly frustrated or express minor complaints about a feature element. E.g., Suggested Feature. Minor or infrequent complaint.

Table - Evaluation Findings (sorted by severity level)

Findings, Evidence, Recommendations, Fixes	***Severity***	***UI Area***	***Round***	***Solution***
Participant didn't know that was TraveLENS, what could he do with it.	1	Tutorial	1	Added tutorial
Participant didn't understand the indicators.	1	Bar Charts	1	Added explanations to indicators
Participant was confused by the twisted bar chart around the circle, it feels like they could be compared between each other but it was quite hard to do that.	2	Choropleth	1	Used parallel coordinates instead of bar charts

Participants thought that some axes seemed to have an inverted relationship to the heatmap (high measure values show lighter, not darker color values).	2	Choropleth	3	Corrected the mistake, now high measure values show darker color values
It was hard for the participant to realize that the top-level item in the drawer was a rollup view.	2	Drawer	3	Added indent to indicate tree structure; Added home icon to the top-level item; Added breadcrumbs in the title bar
Participant could not see blue tooltip for the top level item.	2	Drawer	3	Added tooltip for the top-level item
Participant wanted to select 8-10 countries at a time to compare.	2	General	4	Added the exact number of each indicator
Participant found out that the order of countries was arbitrary.	2	General	4	Ordered the countries alphabetically
Participant wanted to unselect a country.	2	Choropleth	4	Supported unselecting countries
Participant thought that the unselected lines were too dark that made it hard to see the trend and the multi-selected lines.	2	Parallel Coordinates	4	Lightened the unselected lines
Participant thought it was hard to tell which selected country was which color.	2	Parallel Coordinates	4	Used different colors for different selected countries
Participant initially tried to drag in the choropleth legend, thinking it was a control	3	Choropleth	3	
User had initial trouble recognizing the subordination of category/subcategory in the parallel coordinates and scatterplot.	3	Drawer	3	

Participant found there was no country name shown on the map.	3	Choropleth	1	Added country names on the map
Participant preferred to have the average number shown for each indicator as a standard line.	3	Bar Charts	1	Added an average line of the parallel coordinates
Participant preferred to have the exact number of each indicator shown.	3	Bar Charts	1	Added the exact number of each indicator
Participants thought that language could be a big problem when travelling abroad. A index that could measure the ease of communication in specific language is beneficial.	3	Country Detail	2	Added language information of each countries
Participants thought that dive-in crime data in cities or regions can be useful when choosing hotels.	3	Drawer	2	
Participant didn't see that the choropleth WAS a heatmap.	3	Choropleth	3	
Participant didn't immediately see that the heatmap was showing aggregate rankings for categories.	3	Choropleth	3	Added illustration "Aggregates score" under the legend
Participant wanted to see units expressed on the heatmap legend.	3	Choropleth	3	
Basically, it seemed that grayscale made it more difficult for the user to read the choropleth as a heatmap. Participant read the map as though the value shading was there to help distinguish between country boundaries.	3	Choropleth	3	
Participant wanted to see the number of countries selected.	3	General	4	Added the number of countries selected at the top of the country list

Participant could not select new zealand because it was covered by the mini map.	3	Choropleth	4	
Participant thought that it was hard to choose a country from parallel coordinates.	3	Parallel Coordinates	4	
Participant wanted to click on an item which was currently selected to close the drawer.	3	Parallel Coordinates	4	Changed the interaction: the drawer will be closed when clicking on any of the items of the drawer
Participant wanted to reorder the selected subcategories.	3	Parallel Coordinates	4	
Participant wanted to delete some axes that he was not interested in.	3	Parallel Coordinates	4	
Participants thought that different goals of travel led to different information needs. For example, business travel requires individuals to care less about safety because the problems have been covered by companies. While personal travel requires individuals care more on personal safety issues. The goals of the place, whether it is for enjoying the humanities, natural resources or just relaxing on the beach leads to different emphasis on specific kinds of information.	4	Drawer	2	
Participants thought that different regions have different emphasis on regional-interesting data. For example, travelling to southeast Asia may require more information on climates and natural catastrophes, while this kinds of information might not be as important as in travelling to Africa.	4	Choropleth	2	
Participants thought that with data alone, it could be hard to get the tangible information as for how this could be utilized in practice. Suggestions provided based on data	4	General	2	

analysis or giving more tangible examples could be helpful.Insights: providing details-on-demand, or combining datasets at abstract concepts (e.g. CPI) with concrete examples dataset (price of beer).				
Participant didn't immediately see that the heatmap was changing based on subcategory selection.	4	Choropleth	3	
Participant thought the country detail window took up too much space.	4	Country Detail	3	Reduced the space for the country detail window, added Selected Subcategories window below the line chart.
Participant found there was a word/sentence break problem.	4	Tutorial	4	Corrected the word break
Participant found the typo "Megadata".	4	Tutorial	4	Corrected the typo
Participant didn't understand what did selecting parallel coordinates mean.	4	Tutorial	4	
Participant found that the multi-selection in parallel coordinates didn't show up.	4	Parallel Coordinates	4	
Participant thought that an interactive tutorial where user could actually try that specific function out during the tutorial might be better.	4	Tutorial	4	
Participant thought that the unhighlighted parts were too dark, the highlighted area popped up too much.	4	Tutorial	4	
Participant didn't understand what was ICT.	4	General	4	
Participant wanted to be able to rename tabs or see what that tab was about (for example which categories are selected).	4	General	4	

Participant wanted to be able to export the data.	4	General	4	
Participant wanted to see all the other plots brushed and linked when highlighting a line of the legend.	4	Choropleth	4	
Participant wanted to have an interactive legend where he could filter out countries by selecting directly on the legend.	4	Parallel Coordinates	4	
Participant wanted to have a scatter plot with axes indicate two indicators for the parallel coordinates.	4	Scatterplot	4	
Participant thought that the font size was too small.	4	Parallel Coordinates	4	
Participant selected multiple countries, but she didn't know she could change the country of which the tourist arrival data she was looking at by clicking on the arrows.	4	Line Chart	4	Removed the title bar of the line chart, made it part of the country detail information

Appendix C (user research materials)

TraveLens initial survey questionnaire
Survey response summary