Seth Flaxman, Jeffrey Huang, John Stephenson and Xavier Comtesse

CityRank

A dynamic tool for exploring and generating new indices of cities

Keywords: world city, city, ranking, indicator, personal statistics, user-generated

In the context of data on cities, we present an example of how to make statistics relevant and meaningful to non-expert users. While the cities of the world are emerging as key players in global processes, from climate change to migration, the body of data on the cities of the world is neither extensive nor well-organized. Towards the end of organizing, understanding, and presenting this data, we have created an online framework called CityRank. To make this data relevant to users, CityRank allows users to upload new data sets and create and share personalized rankings of cities based on the data included in CityRank’s data repository.

We focus on cities as a compelling and challenging case from the point of view of making data on them interesting and accessible to users. While on the one hand cities are familiar – the majority of the world’s population lives in cities (UN Habitat, 2008, p. 29) – the internal and external processes shaping cities can be complex and hidden from view. The process through which the numerical world of data and statistics is brought into the daily world in which a user feels at home is a central challenge of information design and data visualization. We approach the task of making city data meaningful to users by organizing the data in a simple online interface and by providing a tool for users to create their own rankings of cities.

Why cities?

The study of cities and the networks of which they are a part is a useful analytical lens for studying society over time and in space. Nation states shape the world at an international level, but cities are important and critical actors as well. This has been true historically, before the rise of nation states, and it remains true today. As argued in the literature on world cities, an analysis of the power of a city requires understanding its position and role in relation to other cities. Drawing on social network analysis, world cities can thus be studied as a network of city nodes connected by edges that capture economic, political, or social relationships (Alderson and Beckfield, 2004). As an example, consider the city of Istanbul, which has played a critical role connecting two continents for millennia. To understand this city’s role as a bridge between East and West, one would not start with the modern state of Turkey, which has only come into existence recently. Indeed, Istanbul has been part of various empires over its long history. But what has remained over thousands of years is the city and what has been constant about the city, even as its name and rulers have changed,
is its role as a connector. In the network of the cities of the world, Istanbul has a role to play because of its unique geography, and it continues to play that role as an important world city, whether it happens to be Byzantine, Ottoman, or Turkish.

The United Nations is calling the 21st century the “Century of the City.” In 2008, for the first time in human history, more than 50% of the world’s population was living in cities, whereas urban population represented only 10% of the world population in 1900. By 2050, the world’s population as a whole is projected to be 70% urban (UN Habitat, 2008, p. 29). Cities are where economic activity is organized (Sassen, 2002, p. 17) and a major source of the world’s carbon emissions. They are where we live and where we work. They are worth trying to measure and understand.

**Measuring cities**

The primary way in which researchers try to understand cities is to measure them quantitatively: their attributes, their residents, and their relationships with other cities. There are many dimensions of a city worth measuring, including the following (Alderson and Beckfield, 2004):

- demographics
- societal attributes, e.g. the health and happiness of residents
- economic attributes, e.g. productivity, wealth distribution
- political factors, e.g. transparency, the regulatory environment
- technological factors, e.g. broadband penetration
- relationship with other cities, e.g. air travel connectivity, the number of headquarter and branch offices of multinational corporations.

Cities collect statistics themselves, countries collect data on their cities, and of course NGOs and academics undertake data collection as well. As with any statistical undertaking, statistics on cities are not collected for their own sake but as an instrument towards other ends, be they social, political, economic, or commercial. These ends have a distinguishable effect on what sorts of statistics are collected and how they are made available. In describing city data, we make use of two categories.

- City statistics: quantitative measurements of cities and their residents, such as population, purchasing power, etc.
- Indicators, a term we use interchangeably with rankings: for a set of cities, a preset methodology and underlying statistics are used to calculate an overall score for each city; often these scores are used to assign a rank ordering to the cities.

**Challenges**

We highlight three challenging problems with the current state of data on cities. P. J. Taylor and his colleagues at the Globalization and World Cities group provide a more detailed discussion of these and other problems, which they have been addressing in recent years (Beaverstock et al., 2000). The first problem is in definitions. What is a city? Urban agglomerations, metropolitan areas, and even city boundaries are not precise terms, which means that even such a standard measure as population is not straightforward. This problem is quite critical when considering, for example, creating an index of higher education in which institutions are assigned to cities. This assignment is not always obvious. While Harvard and MIT most likely belong with Boston rather than Cambridge, where do Princeton (with New York City?) or Dartmouth belong? At one extreme, administrative boundaries, however arbitrary, are often used to define cities. At another extreme, Richard Florida et al. have used night-time light emissions to measure mega-regions, rendering even country borders irrelevant (Florida et. al., 2008).
The second problem is in the usability and commensurability of data. Even if cities or organizations have data to share, the task of assembling relevant, accurate data collected using the same methodology on a comprehensive set of the cities of the world is very hard. For this second problem, we believe that with the Internet there is a great, if unrealized, potential for coordination and collaboration, and we have begun addressing this problem through our work on a central repository of city data.

The third problem is with the availability of city indicators. Especially when compared with the abundance of country indicators, there is a marked scarcity of city indicators. There are hundreds of indicators for countries, but only about a dozen for cities. For both countries and cities, these indicators are most useful when they are updated regularly, when their methodologies are transparent, and when the input data sources and final results are freely available. Few city indicators are regularly updated. One caveat is that within specific countries, like the United States, or regions like the European Union, there are more and better indicators, but our focus is on the world, where the situation is quite bleak. Many if not most indicators are created by consulting groups for industry, so often, complete data sets are only available at a prohibitive price for researchers, and even if some data is available to the public, full methodologies are rarely published.

For a few areas of interest, indicators are available, but they have a very specific focus. For example, the consulting group Mercer compiles a Quality of Living Index every year, garnering press headlines for the 50 top cities. As they readily admit, their stated goal is to help “governments and major companies place employees” (Mercer, 2009), not the kind of focus that would be useful for comparing the quality of life in cities for inhabitants in general. The City of London publishes a Global Financial Centres Index (Z/Yen Group, 2009) and MasterCard publishes a Worldwide Centers of Commerce Index (MasterCard, 2008), and both describe their underlying methodologies, a critical step in promoting transparency. There is some academic work as well, like the Digital Governance in Municipalities Worldwide (Holzer and Kim, 2007) and the various academic studies from the Globalization and World Cities Research Network at Loughborough University. But the problem is clear — there are no indicators on cultural production or innovation, intellectual and academic achievements, environmental sustainability, or health of residents.

A framework and a tool

We now describe CityRank, the end result of our research, which can be found online at www.cityrank.ch. CityRank is a framework for organizing city data and a tool for users to create personalized rankings of cities.

Motivation

Our design was motivated by two goals. First, we wanted to explore the problems with city data as outlined in the previous section, share that knowledge with others, and begin thinking about a framework for collaboration to address these problems. Second, we wanted to engage users with the concept of a world city.

In addressing this second goal, we explicitly decided not to create a new, static indicator for cities. We were not in a position to undertake large-scale data collection, so creating a new indicator reflecting an unstudied area related to world city analysis, like soft governance or cultural production, would not have been feasible. Similarly, we decided against creating a new “index of indices” focused on quantifying the concept of a world city because we could not claim authority on the inclusion, exclusion, and weighting of various subcomponents.

Instead, we moved in the direction of a tool which would empower users to interact with data that is
meaningful to them in an engaging way. We recognized the power and attractiveness of indicators, especially conceived as simple rankings from best to worst, eventually settling on the idea of a tool that would allow on-the-fly, user-generated indicators.

While each individual indicator has a very specific, narrow focus and its own biases, our claim is that aggregating them together, if done thoughtfully, will produce a more useful overall ranking. In this way, we still incorporate the idea of an index of indices, while putting decision-making power into the hands of the user. *Allowing users to make decisions for themselves about which factors they value and to what extent makes the underlying data and resulting ranking more meaningful and personal to the user.* For example, one ranking shared by a user of CityRank in the site’s gallery has the title, “Ideal for a scientist.”

**Personas**

In designing CityRank, we considered a few different user personas, which we describe in general terms here. Our first persona is a government employee of a world city, engaged in strategic planning for her city. This user needs data on the competitiveness of her city in relation to other cities. One common task she might want to perform is that of determining the strengths, weaknesses, opportunities, and threats facing her city, commonly known as a SWOT analysis. For this user, CityRank collects various indicators and organizes them by city, thereby allowing a quick comparative analysis of the performance of her city on each indicator. Furthermore, by building a few custom rankings the employee will be able to see which cities rank just above and below her city, which can be interpreted as the opportunities and threats facing her city. And she can use her own expertise to make the criteria and weightings underlying these rankings believable.

Our second persona is an urban planner, engaged in research on global cities with a specific focus on one area, like sports, or culture, or health. This researcher has created a data set measuring his area of interest for a small set of cities. In addition to publishing papers based on his data set, the researcher hopes that other users will show an interest in using his statistics or in helping expand data collection to a larger set of cities. We designed CityRank to aid this process by making it simple for him to share his data on CityRank, and share it with other users who are able to create new rankings of cities based in part on his data.

Our third persona, inspired by a real person, is the head of a trade group in a city. For a number of years, he has been updating a spreadsheet for internal use with thirty or so indicators relevant to his industry, collected for a dozen countries. While country indicators are relevant to him, he is really interested in promoting his city, so he wants CityRank to allow him to transfer his expertise from countries to cities and create a similar spreadsheet comparing cities on indicators relevant to him. And he wants to share the table he builds with CityRank on his group’s website, to further promote his industry in his city.

**The design**

With these personas in mind, we designed CityRank to function on multiple levels, as a data repository, as an interactive tool for building and sharing rankings, and as an open source project which can be adapted to accomplish these goals in different domains.

We collected a variety of city indicators and statistics into a data repository. On the “Cities” page, cities are displayed on an interactive map of the world. Clicking on a city opens a popup with information about the indicators and statistics in which that city appears. On the “Statistics” and “Indicators” (Figure 1) pages the user can
1. Choose Your Indicators

2. Build Your Ranking or upload your own data...

Figure 1. “Indicators” page: allows users to include or exclude indicators to build a ranking in real time, which is shown on the left. Also includes links to a page for each indicator with a visualization of its data and information about its methodology.

browse the statistics and indicators we have collected. From these pages there is a link to a page for each data set. On this page, the user can view the data, read about the methodology, purpose, and source, and visualize the data set on a map of the world. There is also a prominent link inviting users to upload their own data sets as a comma-separated values file. Users must include contact information and cite the source of their data. The system processes the data, which then must be approved by the CityRank team before it is made public. Once it is public, the data is fully part of the data repository. If the data set is of a city indicator, users can build new user rankings based on it.

The tool for building rankings of cities is at the heart of CityRank and it is integrated with the data repository. As the user builds the ranking, it appears on the left sidebar of every page on the site, as seen on the “Indicators” page (Figure 1). On this page a user clicks on checkboxes to include or exclude desired indicators, updating the ranking of cities in real time. From the “Indicators” page users can follow links to pages for each indicator as described above and judge for themselves whether the indicator is one they consider valuable. From the page for a specific indicator, users can also decide whether to include or exclude it, and see how this choice affects their ranking in real time.
After choosing the indicators, the final step of building a ranking occurs on the “My Ranking” page (Figure 2). On this page, users drag sliders to assign weights to each indicator, expressing their preferences by assigning a value of “Importance” on a scale from one to ten.

Just as on the “Indicators” page, when a user adjusts the slider, the ranking on the left sidebar is updated in real time. If users are so inclined, they can share their rankings in the “Gallery.” Users are asked to give their rankings a title and a description, and rankings are represented visually using a custom-generated icon based on the user’s choice of indicators and weights. Although we were not sure in advance whether users would be inclined to share their rankings, we thought that including the capability was important as a proof of concept. Our experiences on this front are discussed in the final section.

The choices a user makes about including and weighting indicators, combined with the ability to upload new data sets, are central features of CityRank’s design. It is important to remember that we do not claim that the rankings CityRank creates have authority in the same way that a published indicator claims authority. However, this claim of authority can be challenged: existing rankings have their own biases and shortcomings, as we described earlier in the section on measuring cities.

When a user includes an indicator, she is implicitly including the assumptions underlying that indicator. In our ideal world, users would have full access to the data and methodology used to generate these indicators, so that users could more finely tune and personalize rankings of, for example, quality of life or price and earnings before including these rankings in overall rankings. As a step towards addressing this problem, informed users
can assign lower values of importance to indicators they value or trust less. Highly motivated users can also upload their own data sets for inclusion in their final rankings. It is through the use of these features that CityRank allows users to make informed decisions and create personalized rankings, not authoritative, but valuable nonetheless.

The algorithm

As we described earlier, the underlying methodology by which an indicator is generated is critically important, so we chose an appropriate algorithm carefully. Since we were interested in creating a new ranking from a variety of existing incommensurable indicators, we needed to formalize the idea of combining different indicators and to find a suitable algorithm to do this combination. Our inputs and desired outputs were most closely matched by a problem known in computer science as the rank aggregation problem, and the recent publication of a new algorithm which could handle problems of our type convinced us that this approach was the most sensible for our purposes. The algorithm is due to Jiang et al. (2008). We include a high-level description of the algorithm here because it is central to the functioning of CityRank and because we would like to promote its use in other contexts.

Table 1 shows a stylized instance of the rank aggregation problem: four different made-up indicators compare New York, Geneva, London, and Chicago. This instance reflects a real problem with indicators: some of the indicators do not include all of the cities. The rank aggregation problem is simple: find a single output ranking which is as close as possible to each of the original rankings.

What makes rank aggregation non-trivial and interesting is that indicators can, and usually are, in direct conflict: in Table 1, Indicator A ranks New York higher than Geneva, which is in turn ranked higher than London. But Indicator B ranks London higher than New York, and does not even include Geneva. One way to visualize the rank aggregation algorithm is to create a pairwise comparison graph for the indicators. Each city is represented on the graph by a vertex, and a directed edge is included between every pair of cities, with the direction determined by the ranking, from higher ranked to lower ranked. The first indicator in Table 1 would be represented by the graph in Figure 3. Adding the edges implied by Indicator B to the graph leads to conflicting edges and cycles, as shown in Figure 4. We add weights to each edge based on the differences between pairs of scores in Table 1, e.g. for Indicator A, New York scores 100 and London scores 25, so the weight on the edge from New York to London (this edge was omitted from the figure for clarity) would be 75. The problem of rank aggregation can thus be formulated as follows: find a sensible way to create a single graph, without any inconsistencies or cycles, with the objective of making

<table>
<thead>
<tr>
<th>Rank</th>
<th>Indicator A</th>
<th>Indicator B</th>
<th>Indicator C</th>
<th>Indicator D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>New York 100</td>
<td>Chicago 100</td>
<td>Chicago 100</td>
<td>Geneva 100</td>
</tr>
<tr>
<td>2nd</td>
<td>Geneva 50</td>
<td>London 50</td>
<td>New York 50</td>
<td>New York 50</td>
</tr>
<tr>
<td>3rd</td>
<td>Chicago 37.5</td>
<td>New York -62.5</td>
<td>Geneva -125</td>
<td>Chicago 50</td>
</tr>
<tr>
<td>4th</td>
<td>London 25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Four indicators comparing New York, Geneva, Chicago, and London. Ranks are shown on the left, and normalized scores are included next to each city.
the graph as close as possible to the original, that is, without reversing too many edges or changing their weights too much.

The first step of Jiang et al.'s algorithm is to consider the pairwise comparison graph for each input indicator. The algorithm aggregates these graphs into one graph, the aggregated graph, with each edge weight calculated as a weighted average of that edge over all of the graphs which contain it. This is illustrated in Figure 5. The aggregated graph may still have cycles, as in the New York, Geneva, London cycle in Figure 5, so the final step is to perform a least squared minimization over the space of pairwise comparison graphs (with least squared distances calculated to the aggregated graph) to search for a similar graph, but without any cycles. In other words, the algorithm tries to find a graph as close as possible to the comparison graph by changing some of the pairwise comparisons between pairs of cities, either by reversing the arrows altogether (flipping their order in the ranking) or just by changing their weights slightly.

**Discussion and potential future adaptations**

The design of CityRank draws on ideas from direct manipulation, search engines, and user-created lists. With “The Attribute Explorer,” Tweedie et al. (1994) gave an early example of the now ubiquitous interface in which a user filters a set of possible results using direct manipulation of a fixed set of criteria. “The Attribute Explorer” was an apartment search tool in which users filtered apartment listings by a variety of relevant factors until they arrived at a manageable small set of results.

In a related but separate setting, there are myriad approaches to the design of search engine interfaces, from strict regular expression searches to natural-language search. There are also varying designs for search engine results, from the now-standard list ordered by relevance (as typified by Google), to grids and 3-D interfaces. Most search engines return results in a list ranked by relevance, where relevance can be defined in various ways. In both search engines and direct manipulation
tools, the focus is on helping a user find a result from a (possibly huge) set of possibilities.

In a third setting, many websites allow users to create and share their own rankings, such as Amazon’s “Listmania!” (top ten style) lists. Here, the focus is on allowing users the freedom to create and share any ranking they desire with others.

Our approach is innovative insofar as it blends all three approaches. In creating a ranking the user is able to directly manipulate the ranking through the inclusion and exclusion of sub-indicators and the sliders through which users assign importance to each sub-indicator. The ranking presented to the user can be thought of as being ordered by “relevance” according to the user’s personal criteria of results in an analogy to search. Finally, the end result for a user is a ranking of cities, akin to a user-generated top ten list. We hoped that the user would not focus merely on the creation of a single final ranking, but would use the tool to explore the cities and indicators themselves, to discover the strengths and weaknesses of the cities and question the assumptions and methodologies of existing indicators.

CityRank was designed as a way for expert and non-expert users alike to explore existing city data and think about what types of new city data would be useful in the future. In informal trials, non-expert users have found the idea of building their own ranking compelling. However, without guidance, some users have been confused by parts of the interface, especially the aggregated ranking in the left sidebar which is present on every page (see Figure 2). On pages for specific indicators or statistics, users became confused about what the aggregated ranking had to do with the specific data set they were viewing (depending on the situation, it might have been included as a sub-indicator or it might have been totally independent).

An interesting situation which occurred in informal testing was one in which users first selected an indicator based on its short description, for example, “Power.” Later, when assigning a weight to this indicator or simply seeing the cities which ranked high on it, users wanted to know what this indicator meant, what its criteria and methodology were (it is taken from Alderson and Beckfield, 2004). Though the answer to this important question could be found on the website, users needed to be actively guided to the correct page by one of the developers. A simple way to resolve this issue would be to add a short amount of explanatory text on the “Indicators” page with an obvious link to the indicator page which contains the data, more information, and a link to the original source. This simple change would encourage users to ask more questions about the data sources, questions CityRank is able to answer to an extent, thereby encouraging them to make more informed choices about which indicators they value and to what extent. Adding a comments or ratings feature would also allow users to share their appraisals of certain indicators with others.

Experts and non-experts alike found the coverage of indicators lacking. Many users, for example, wished that they could select indicators on the environment, civil society, and media. As we described in the section on city data, these indicators simply do not exist at present for cities. We would like to promote the idea that these are interesting indicators for researchers to create in the future, but this goal was not apparent to a casual visitor to CityRank, so some users were led to believe that this was an oversight on our part. Users were also quick to point out the shortcomings of various indicators, and as we described in the design section, in an ideal world users could remix and reweight the data underlying these indicators as well.

A common theme in a number of presentations at the Data Designed for Decisions conference (Paris, 2009) was the desire and ongoing search for ways to make statistics more relevant and personal to users. We hope that CityRank can serve as an example of one attempt to encourage users to interact and play with data that is or can become meaningful to them.
As an example of another domain where this idea could be compelling, we have recently launched a site based on CityRank called UniRank (online at www.unirank.ch), which allows users to build university rankings based on a variety of indicators. In contrast to data on cities, there is a wealth of relevant and current data on universities. One major problem with existing university rankings is that they reflect the choices and beliefs of their creators about what makes a good institution. As we have argued at length, opening the methodology up to the user addresses this problem while engaging the user in the underlying data, which is what is most relevant. We recently became aware of other work along these lines: Geoff Davis's PhDs.org (www.phds.org) is a compelling and useful tool for creating personalized rankings of graduate programs in the United States.

While user-generated rankings could be seen as undermining the authority of rankings created by experts, there is no reason both expert and personal rankings cannot coexist. For example, the Organisation for Economic Co-operation and Development (OECD) team responsible for the Social Institutions and Gender Index of countries (online at www.genderindex.org) is currently adding the ability for users to experiment with their published indicator by choosing and weighting its component sub-indicators. The project will be at my.genderindex.org and it will use the code underlying CityRank.

For those interested in extending or experimenting with the ideas of CityRank, the source code is available as an open source project online at www.cityrank.ch/code. The code has been generalized to smoothly handle cities, countries, universities, etc. CityRank was developed in Python with the Django framework, and it uses Google's Data Visualization API and the Yahoo! User Interface Library. The rank aggregation algorithm was implemented in Python and C using Manolis Lourakis's levmar library (online at www.ics.forth.gr/~lourakis/levmar) for calculating the least squares optimization. Jiang et al.'s rank aggregation algorithm could easily be replaced by a more appropriate or simpler algorithm, depending on the setting.

Our work focused on information and algorithm design, and there is much work yet to be done on the data visualization front. Many geographic data visualizations have become ubiquitous, from maps of the world with land masses scaled to represent a variable, to red, blue, and purple maps of the United States showing electoral trends. However, there is at present no compelling map of the world from the point of view of its cities, and no tools providing a simple and sensible way to visualize city indicators, much less data on the relationships between cities. This could be a very promising direction for future research.

Notes

1. In practice, indicators from different data sources are scaled differently, so they must be normalized before they can be aggregated. We observe that these indicators include rankings, so we can normalize the raw scores based on their ordinal scores (i.e. ranks). In effect, this means that all cities which rank first are scaled to have the same score (100), all cities which are ranked, say, 25th are scaled to have the same score (50), and all other scores for each indicator are scaled accordingly (we use a linear scaling, with a line determined by the original and scaled scores of the cities that rank first and 25th).

References


About the authors

Seth Flaxman received his bachelor’s degree from Harvard University in computer science and mathematics. He worked as a research assistant in the Media and Design Laboratory at EPFL for one year. He is currently working on a PhD in computer science at EPFL.

Jeffrey Huang is the Director of the Media and Design Laboratory at EPFL and a Full Professor at the Faculty of Computer and Communication Sciences, and at the Faculty of Architecture, Civil and Environmental Engineering (ENAC). His research explores the vision of fusing physical and virtual architectures. His current research and teaching projects include the examination of emerging architectural typologies (workspaces, banking, museums, theater), novel, organic approaches to architectural design (parametric design, algorithmic design), data-driven visualizations (such as cityrank.ch), the integration of physical computing (sensors, actuators, RFID, LEDs, mobile devices) into architecture and cities, and, more generally, design-centered approaches to human–computer interaction (design thinking).

John Stephenson graduated from the Graduate Institute of International and Development Studies (IHEID) in Geneva. He worked at the Fondation pour Genève and is now with eBay, Inc.

Xavier Comtesse is the head of the Observatoire at the Fondation pour Genève. He has long been reflecting on the implications of new technologies and on ways Geneva can make better use of globalizing forces.

Contact

Seth Flaxman  
Ecole Polytechnique Fédérale de Lausanne  
EPFL IC ISIM LDM1 BC 101 (Bâtiment BC) Station 14  
CH-1015 Lausanne, Switzerland  
seth.flaxman@epfl.ch

Jeffrey Huang  
Ecole Polytechnique Fédérale de Lausanne  
EPFL IC ISIM LDM1 BC 101 (Bâtiment BC) Station 14  
CH-1015 Lausanne, Switzerland  
jeffrey.huang@epfl.ch

John Stephenson  
eBay, Inc.  
john.stephenson@graduateinstitute.ch

Xavier Comtesse  
Fondation pour Genève  
Domaine La Pastorale 106, route de Ferney  
CH-1202 Genève, Switzerland  
xavier.comtesse@avenir-suisse.ch