YouthfulCities Index

Responsive Attribute Weighting and Interactive Visualization

**Kie Gouveia**

**October 21, 2015**

# Introduction

This project will entail analysis related to composite indices, pattern mining, and visualization, and will leverage the YouthfulCities Index dataset provided by YouthfulCities. During the initial briefing meeting, it became evident that the YouthfulCities team desired a way to unpack which indicators are necessary to rank youthfulness as well as a logical method to weight each category in order to calculate relative ranks. The goal of this project will be to support these two objectives by assessing/modifying the current weighting system and by allowing individuals to weight indicators according to their own specific needs and interests.

In addition to allowing for interactivity and personalized data exploration, the tool will seek to engage users visually through dynamic, responsive visualizations which will be generated using RShiny.

# Literature Review

In order to prepare for this project, I reviewed sources dealing primarily with two topics:

1. The development of composite indices (specifically pertaining to the ranking of cities, where possible).
2. Data visualization techniques

Topic 1 will allows for proper assessment and development of a ranking algorithm to compare cities, while topic 2 will provide the foundations required for creating a highly interpretable data visualizations.

**Handbook on Constructing Composite Indicators – Methodology and User Guide**[**1**](http://www.oecd.org/std/42495745.pdf)

**Organization for Economic Co-operation and Development**

This handbook provides a complete guide to constructing and utilizing composite indicators. Specifically, the book is concerned with composite indicators that are designed to compare and rank country performance. By outlining technical guidelines and best practices for formulating and leveraging composite indicators, this handbook will be an indispensable resource for investigating and potentially optimizing the indices used by Youthful Cities. It will also provide insight into the development of the algorithm which will leverage user preferences to re-weight characteristic according to personal interests, needs, and preferences.

**Composite Indicators of Country Performance: A Critical Assessment**[**2**](http://www.oecd-ilibrary.org/docserver/download/5lgsjhvj7lbt.pdf?expires=1445386908&id=id&accname=guest&checksum=48CDAEF037DAFC837CBBBC0C60CB1E37)

**OECD Science, Technology and Industry Working Papers**

This paper provides a robust overview of the methodological difficulties associated with the development of composite indicators, specifically those being used to rank countries. Along with simply reviewing the challenges associated with creating composite indicators, this paper also provides steps in constructing composite indicators and makes suggestions to ensure transparency and appropriate use of indices in terms of analytics and policy.

**Grammar of Graphics**

**By: Leland Wilkinson**

This book is the basis for the widely used ggplot2 visualization package in R. This book is designed to provide a strong foundation for producing highly quality, easily interpretable quantitative graphics. Essentially, the book is designed to create a formal system of rules for generating graphics. Given that my project will primarily focus on data visualization, this book will serve as an indispensable reference guide as I work to make the final as useful as possible.

**Graphical Perception: theory, Experimentation and Application to the Development of Graphical Methods**[**3**](http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/xpls/abs_all.jsp?arnumber=6327269&tag=1)

**Journal of the American Statistical Association**

**By: William S. Cleveland; Robert McGill**

This is a foundational paper in the field of statistical graphics. It outlines how characteristics such as size, colour and length compare in terms of visually portraying quantitative information. In creating an interactive, visual representation of the data, it will be imperative to take such considerations into account to minimize the chances of misinterpretation.

**Graphical Overlays: Using Layered Elements to Aid Chart Reading**[4](http://vis.berkeley.edu/papers/grover/)

**UC Berkeley**

**Nicholas Kong; Maneesh Agrawala**

This paper specifically details graphical overlays which can be used to augment charts by overlaying additional elements to support chart interpretability. It identifies five main types of overlays that help to support perceptual and cognitive processes. Overlays such as these may allow for improved interpretability and for more pertinent information to be conveyed to users on a single screen.

# Dataset

2015 YouthfulCities Index Ranking

According to youthfulcities.com, the YouthfulCities Index is an ambitious collaborative effort to analyze the cities around the world from a youth perspective. It is designed to look at how youth live work and play in their urban setting, with the objective being to decipher how youth can become more engaged in their cities.

Data for the most recent iteration of this index was collected between September 2014 and April 2015, but primarily reflect information that is current as of 2013 – 2014. Data was collected in 55 cities around the world. The index is a composite index made up of 20 attributes, which are further defined and measured based upon 101 indicators, listed below:

|  |  |
| --- | --- |
| **Safety** | **Age standardized road deaths** |
|  | **Food safety** |
|  | **Age standardized suicides** |
|  | **Homicides** |
| **Affordability** | **Consumption tax rate** |
|  | **Rental housing** |
|  | **GINI Coefficient** |
|  | **Cost of one prepaid cellular minute** |
|  | **Food - 12 eggs** |
|  | **Monthly transit pass** |
|  | **Price per square meter to buy an apartment** |
|  | **Tube of toothpaste** |
| **Transit** | **KMs of dedicated public transportation** |
|  | **HRs/week dedicated public transit** |
|  | **Public bike rentals** |
|  | **KMs of bike paths** |
|  | **Commuter time by transit** |
|  | **Commuter time by foot** |
|  | **Commuter time to airport** |
|  | **Real time transit app** |
|  | **Density** |
|  | **Standing rate for a taxi** |
| **Health** | **Number of public health clinics** |
|  | **Number of public sexual health clinics** |
|  | **Number of homeless shelters** |
|  | **Urban smoking scale** |
|  | **Mental health outpatient facilities** |
|  | **Doctor density** |
| **Travel** | **Number of cities connected through direct flights** |
|  | **Getaway city bus cost** |
|  | **Getaway city bus time** |
|  | **Getaway city bus frequency** |
|  | **Getaway city plane cost** |
|  | **Getaway city plane time** |
|  | **Getaway city plane frequency** |
|  | **Hostel Stay** |
| **Employment** | **Youth unemployment rate** |
|  | **Employment initiatives and programs scale** |
|  | **Youth Employment Centres** |
|  | **Number of new jobs created in 2013** |
| **Environment** | **Municipal water scale** |
|  | **Smart cities initiatives scale** |
|  | **Quantity of annual recycled materials** |
|  | **Quantity of annual waste** |
|  | **Carbon emissions** |
|  | **Recycled materials** |
|  | **Total registered vehicles** |
| **Education** | **Post-secondary institutions** |
|  | **Average cost of tuition** |
|  | **Full-time undergrad enrolment** |
|  | **Student housing** |
|  | **Student debt** |
| **Entrepreneurship** | **Age at which you can register a business** |
|  | **Number of entrepreneurship Incubators** |
|  | **Total early stage entrepreneurial activity** |
|  | **Ease of doing business rank**  **Employment initiatives and programs scale** |
| **Public space** | **Total green space/public space** |
|  | **Number of municipally maintained sports facilities/fields** |
|  | **Number of public libraries** |
| **Financial services** | **Business banking availability** |
|  | **Personal Banking availability** |
|  | **Number of "Chartered" banks in your city** |
|  | **Online banking** |
|  | **Mobile banking** |
|  | **Financial literacy** |
| **Diversity** | **Languages to vote in** |
|  | **Diversity of food** |
|  | **Openness to LGBTQ** |
|  | **Openness to immigrants** |
|  | **Diversity of religion** |
|  | **Global gender gap index** |
| **Digital Access** | **Cellular competitiveness level** |
|  | **Average cost of baseline package with data** |
|  | **Free WIFI scale** |
|  | **Mobile phone infrastructure** |
|  | **Municipal government open urban data scale** |
|  | **Gamers + Developers scale** |
| **Music** | **Number of nightclubs** |
|  | **Music festivals** |
| **Creative arts** | **Grafitti and street art** |
| **Sports** | **Number of professional sports teams** |
|  | **Number of pro sports facilities** |
|  | **1 pair of sport shoes (Nike, Adidas, or similar)** |
|  | **1 month of gym membership in business district** |
| **Film** | **FIlm festivals** |
|  | **Cost of a movie ticket** |
|  | **Number of cinema seats** |
| **Civic engagement** | **Voting age** |
|  | **Average age of city councillors (elected officials) in your city** |
|  | **Volunteer opportunities** |
|  | **Political influence - Youth advisory board** |
|  | **Volunteerism in high school scale** |
| **Food and nightlife** | **Number of restaurants in your city** |
|  | **Last call index** |
|  | **Number of food festivals** |
|  | **Food - fast food** |
| **Fashion** | **Young designer showcase scale** |
|  | **Number of fashion incubators** |
|  | **Number of design schools** |
|  |  |

The methodology for the index is comprised of two steps:

1. **Determining what to measure in each city.** This is accomplished through a separate survey, called the Urban Attitude Survey. This tool is used to determine what characteristic are important to youth in the cities they live in**.**
2. **Collecting data to measure cities:**  Leveraging the information garnered from the Urban Attitudes Survey, primary and secondary data is collected to measures 20 attitudes and 101 indicators which we deemed important to youth. This data, along with weights determined by relative importance of each indicator, are used to develop a ranking of cities.

For information regarding data processing, including normalization, comparing different years, scale and boundary issues, imputation of missing values, and data credibility, the YouthfulCities webpage offer more information [here](http://www.youthfulcities.com/#!youthfulcities-index/c10wk).

# Approach

## Step 1: Clean data

The first step will be to do some preliminary cleaning and recoding within the Youthful Cities excel spreadsheet. For example, there are a few broken equation chains which must be fixed in excel prior to being loaded in R.

Once very preliminary cleaning is done, the data will be imported into R where the data will be restructured and further cleaned to allow for easy access and development.

## Step 2: Assess current ranking algorithm

Once the data has been transformed into a usable format, it will be possible to scrutinize the the current weighting system further. Initial steps will include vetting indicators to ensure that each variable is contributing new information and that the weighting algorithm is producing an unbiased ranking of the cities. Best practices will be garnered from the research papers listed above.

## Step 3: Modify algorithm and account for user input

Depending on the findings in Step 2, the algorithm may be modified. Regardless, an algorithm will be developed which will allow for user preference to be taken into account while ranking the cities.

## Step 4: Develop and implement interactive visualization

Finally, an interactive visualization will be developed in R Shiny which will allow users to select attributes and indicators that are important to them. Once the cities have been re-ranked, taking user preferences into consideration the tool will display top/relevant cities in a graphical representation (potentially a map format).