Group 1 Information Visualization Report

| Tiago Nascimento  Instituto Superior Técnico  Alameda  70493 | Miguel Cruz  Instituto Superior Técnico  Alameda  76102 | Daniel Trindade  Instituto Superior Técnico  Alameda  76349 |
| --- | --- | --- |

# INTRODUCTION

How do the countries that participated in the Olympic Games stand against each other concerning the medals they achieved through the years? Do countries with a greater population also get more medals? How do these standings evolve over time and how do they accumulate in a certain amount of years?

We knew there was data to answer how many medals a country scored for a certain sport in a certain year, and we did find a solution that did that, on the Internet. But we wanted to go a bit further and be able to make comparisons, not just for one sport, not just for one year at a time, and not just counting one or all kinds of medals. So we went further and now we could know, as an example, if Russia had more or less gold and silver medals than the Soviet Union.

We also thought of seeing how many medals each country “owned”. That is, for example, how many medals Germany had scored, plus medals Germans playing for other teams scored, minus the ones foreigners playing for Germany scored. Unfortunately, we couldn’t find the nationalities of a big amount of athletes, so we decided to leave this feature alone.

The first tasks we proposed to support were, then:

* Browse – display the countries with the most gold medalists in total in a given year.
* Identify – show the country with the most medalists in a sport of all time.
* Locate – show the position of a country in the overall standings.
* Explore – using the coefficient medals/population (derivative variable), display the countries with the highest coefficient.
* Compare – show the medals each country won.

Our initial thought of showing statistics for “all time” was also changed to a “span of years”, where we chose the minimum and maximum years, making our visualization more flexible.

# The data

After days of search on the Internet, we found several documents/datasets that provided to us some contents of what we were looking for. For the main problem, that was giving to us data, divided by section of years/Olympic edition, about the medals and the countries that have won those medals in different sports, we use a dataset that we’ve found in this article of the Guardian: <http://www.theguardian.com/sport/datablog/2012/jun/25/olympic-medal-winner-list-data#data>. Later, if we wanted the original dataset, this site would redirect us to this doc - <https://docs.google.com/spreadsheets/d/1zeeZQzFoHE2j_ZrqDkVJK9eF7OH1yvg75c8S-aBcxaU/edit>. This was the original dataset where we based our data.

To represent a bubble chart of the country in a world map we needed another dataset that represented the location of a country in coordinates so we could associate a country in a map image that would be properly related with the attributes of size image and GPS coordinates. For this, we opted to use this dataset: <https://developers.google.com/public-data/docs/canonical/countries_csv>.

Finally, we also want it to test if a countries population had some impact in order of winning more or less medals in an edition of the Olympic Games. So for that we only concerned in data that had the population of all countries and we found a dataset with that information, but it only had information from the year 1960 to now: <http://data.worldbank.org/indicator/SP.POP.TOTL?page=6>.

After founding and extracting this data, we had to evaluate them in order to know if we could relate attributes between them; like for example, could we say in the bubble chart that Romania is located in such coordinates given in the GPS dataset? No. Could we associate the Germany population in 1960 with the winning rank they had in that year? No. And so on. There were things that we had to have concerned like the ISO, NOC and IOC code, the countries that no longer exists but existed in the pass (West Germany, Soviet Union for example) and some information that we just didn’t need it.

So for the first issue, we had some problems like ISO 2-letter codes and ISO 3-letter codes, and for that, we made a dataset relating those codes. Also we had problems cause there were codes that in time had changed, so for that we put in the original dataset the recent ISO code in all occurrences of the country (ies). We also had a problem in countries that no longer existed. Like, what would be the coordinates for West Germany? So for that, we add to the GPS coordinate dataset an ISO code that represented the older country, its name and a latitude & longitude that were located inside of the corresponding countries that they were located. Meanwhile, when we were doing this task, we had the concerned to grant to the Olympic dataset (the data that had all winners and medals) the corresponding ISO and NOC code updated of countries that don’t exist anymore and in the population dataset, we incorporate the same task.

The third issue that we talked about it, was the unused information that we had on some dataset. Like in the Olympic dataset, we had attributes like City, Athlete, Gender, Event and Event\_gender that we didn’t need it. So we extract that information from the data.

In the end, we had 2 dataset. One containing the Olympic medals (each year, we had information about a sport, the country, and its ISO 3 and 2-letter code, that win the medal of that sport, and the medal that it won) and the coordinates of the country, and a dataset of the population over the years that could be related with the earlier dataset by associating the ISO-3-letter codes with the NOC codes.

# VISUALIZATION

## Overall Description

The group agreed that for this visualization, it would be necessary 3 tabs. One for Standings, one for Standing Comparison and last, but not least, one for Coefficient. All of them were created to make easy for the user to search/explore results and data. To do so, we had several inputs that could be modified to the user needs and wishes. This inputs are the measures that will determinate the visualizations that we called, the output. There is in all tabs 3 ways of visualization/outputs that show the data that the user pre-defines in the inputs.

The first tab (Standings) was developed with a left vertical sidebar with all Olympic sports, so the user can select the sport is interested on searching, 3 boxes that represent the medals he wants to lookup, a horizontal sidebar bellow the boxes that can be used so that an user can explore standings in a specific year or an interval of years of his choice. Bellow this sidebar, we have a “search for a country:” text box, where the user can search for a specific country. All of this were attributes of searching, i.e. the input that the user can provide to the system. The visualization, our output, is represented by 3 visualizations: a map bubble chart, where a country is represented in the map with a bubble (the bigger the standing, the bigger the bubble would be), a rank bar chart, where the standings of the countries would be presented in a rank type, in which the length of a bar that was bigger, represented the country with the most input options given by an user, and the next bars would be also represented the same way, but in a decreasing way, just like a ranking. These bars would have, after the bar, the number of medals a country won, and before the bar, the correspondent country. There was still a visualization – a line chart - that the group made to know the standings of a determinate country given by the user to know its standings over the years.

The second tab, the standings comparison, had the same attributes that the first tab had, but this time, it hadn’t the left vertical sidebar with the sports because all sports could be compared in the bar chart that this tab would have. Another particularity is that this tab has two search text boxes, so the user can compare the standings between 2 countries. As for the output, we would still have the map bubble chart, a double bar chart (this time, it wouldn’t be ranked because it would be centered in the sports; each country had its color) and a line chart with two lines that represent the user selected countries performance over the years (each country had its color).

In the third tab, Coefficient, we represent the impact the population had on the winning medals of all countries in a year. For that, we had only 2 inputs: a horizontal sidebar to select a year, and a search box to lookup for a certain country. The outputs/visualization were a map bubble chart like we used in the other 2 tabs, a rank bar chart like we used in tab 1, and a line chart of the country coefficient medal/population over the years.

**Rationale**

Our choices were based on the simplicity of the visualizations. We wanted to grant outputs that make the user understand all the information. For that we though immediately on label texting to orient the viewer of our visualizations. But of course, some visualizations weren’t compatible with text labelling, i.e., mixing text labelling with some charts weren’t very appropriate. Like for example, in the bubble chart, we originally thought of granting inside the bubble a text label of the amount of medals the bubble of a country had; but that wasn’t a good idea because if a bubble was small, it would be hard to see the label. So we thought of another idea, which was: when a user passes with the mouse on any bubble, bar or point of a line chart, it would pop up a window with information about that point, for example, when we are passing our mouse in the bubble chart in the bubble of United States, the window that would pop up would be “USA – 1052 medals – United States”. For the bar chart the window would look the same. For the line chart, it would appear the amount of medals in a certain year (of course the text is different according to the tab the user is).

The map bubble chart was used to make the user comfortable in locating a country. If he preferred and if the country had a proper sizer bubble (if it was visible) then we can press in the bubble, and locate in all the other visualizations the selected country that would have a different color from the other countries. This is one alternative to the search box in the tab. But if a country hasn’t have a good size to be located or the user doesn’t know where a certain country is, its recommended to use the search box. This is only an amuse, cultural and direct way of viewing the data in a World Map.

For the bar chart we gave labels after and before the bar so the user would know what the bar represented. The before label would declare the country the bar is representing (in the second tab, the sport), and the after labelling would declare the size of the bar (the length of medals or coefficient population/medals in a country). In the tab 1 and 3 we used it in a rank type way, so we could give to the user the idea of what is the country or countries with high standings in a predefined input of the user. After all, what user wants from this is statistics to know which were the best countries on a certain input (a certain problem/task). We used this visualization cause its intuitive for the users to understand the results and get results in a quick and effectively way. If the user is searching for a certain country in the search box, it would be shown in a different color the selected country in the bar. Or if the results were there and the user didn’t select in the search box for a country, then he could select a label and the country would be selected. In the second tab, we didn’t make it like a rank type because we wanted to compare statistics between countries and sports. It’s risky in this case because we’re comparing two lengths side by side but since were using labelling to know the “size” of the bar, then it works fine in this case.

Our final visualization – the line chart – was used because we wanted something else than just have visualization of a country in a certain year. We wanted to grant to the users the possibility of evaluating a pre-defined country according to an interval of years, in this case, the standing of a country between 1896 and 2008. This way we had more information to the board and provided to users more information about a country and the input provided by a user.

Simplicity was the original key for our visualizations and we think that we achieved that gold. But as were doing this project, we based also in another key that was knowledge. We were always looking for ways to provide the most information possible to the users and we think that we did a good work, but it could be better; maybe testing with users could give to us some answers on what a user could want from our system and if we had enough information to provide to them. Maybe a future update of the system can happen.

## Demonstrate the Potencial

To be continued…

## Normal or Body Text

Please use 10-point Times New Roman or Times Roman font (the Normal style) or, if this is unavailable, another proportional font with serifs, as close as possible in appearance to Times New Roman 10-point. Other than Helvetica or Arial headings, please use sans-serif or non-proportional fonts only for special purposes, such as source code text.

## First Page Copyright Notice

This sample word document has the correct ACM SIGCHI copyright notice in place (see page 1, bottom of column 1). Accepted papers will be distributed in the conference publications. They will also be placed in the ACM Digital Library, where they will remain accessible to thousands of researchers and practitioners worldwide. ACM’s copyright and permissions policy is here:

<http://acm.org/publications/policies/copyright_policy>

## Subsequent Pages

On pages beyond the first, start at the top of the page and continue in double-column format. The two columns on the last page should be of approximately equal length.



Figure . Use high-resolution images, 300+ dpi, legible if printed in color or black-and-white. Number all figures and include captions below, using Insert, Caption.

## References and Citations

Use a numbered list of references at the end of the article, ordered alphabetically by last name of first author, and referenced by numbers in brackets [1,3,4].

| Objects | **Caption – pre-2002** | **Caption – 2003 and afterwards** |
| --- | --- | --- |
| Tables | Above | Below |
| Figures | Below | Below |

Table . Table captions should be placed below the table. We recommend table lines be 1 point, 25% black. Minimize use of unnecessary table lines.

Your references should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible (i.e., you provide the address for obtaining the report within your citation) and may be obtained by any reader for a nominal fee. Proprietary information may not be cited. Private communications should be acknowledged in the main text, not referenced (e.g., “[Borriello, personal communication]”).

References should be in ACM citation format: <http://acm.org/publications/submissions/latex_style>. This includes citations to internet resources [1,4,8,1] according to ACM format, although it is often appropriate to include URLs directly in the text, as above.

# SECTIONS

The heading of a section should be in Arial 9-point bold, all in capitals (Heading 1 style). Sections should not be numbered.

## Subsections

Headings of subsections should be in Arial 9-point bold with initial letters capitalized (Heading 2 style). For sub-sections and sub-subsections, a word like *the* or *of* is not capitalized unless it is the first word of the heading.

### Sub-subsections

Headings for sub-subsections should be in Arial 9-point italic with initial letters capitalized (Heading 3 style).

# FIGURES/CAPTIONS

Place figures and tables at the top or bottom of the appropriate column or columns, on the same page as the relevant text (see Figure 1). A figure or table may extend across both columns to a maximum width of two columns, or 17.78 cm (7 in.).

Captions should be Times New Roman 9-point bold (Caption style). They should be numbered (e.g., “Table 1” or “Figure 2”), centered, and placed beneath the figure or table. The words “Figure” and “Table” should be spelled out (e.g., “Figure” rather than “Fig.”) wherever they occur.

All figures should also include alt text for improved accessibility. In Word, right click the figure, and select Format Picture | Layout | Alt Text). Papers and notes may use color figures, which are included in the page limit; the figures must be usable when printed in black-and-white in the proceedings.

The paper may be accompanied by a short video figure up to five minutes in length. However, the paper should stand on its own without the video figure, as the video may not be available to everyone who reads the paper.

## Inserting Images

Occasionally MS Word generates larger-than-necessary PDF files when images inserted into the document are manipulated in MS Word. To minimize this problem, use an image editing tool to resize the image at the appropriate printing resolution (usually 300 dpi), and then insert the image into Word using Insert | Picture | From File.



Figure . Sample of a wide figure. Be sure to place at the top or bottom of the page. Ensure that important information is legible in both black-and-white and color printing. Image: CC-BY-ND ayman on Flickr.

## Table Style

The text of tables will format better if you use the Table Text style (as in Table 1). If you do not use this style, then you may want to adjust the vertical spacing of the text in the tables. To adjust the spacing of text in a table in Word, use Home | Paragraph | Indents and Spacing. Generally, text in each field of a table will look better if it has equal amounts of spacing above and below it, as in Table 1. Table captions should be placed below the table. We recommend table lines be 1 point, 25% black. Minimize use of unnecessary table lines.

For improved accessibility, header rows of tables should be marked. In Word, right-click a header row, and select Table Properties | Row | Repeat as header…

# LANGUAGE, STYLE AND CONTENT

The written and spoken language of SIGCHI is English. Spelling and punctuation may use any dialect of English (e.g., British, Canadian, US, etc.) provided this is done consistently. Hyphenation is optional. To ensure suitability for an international audience, please:

* Write in a straightforward style.
* Try to avoid long or complex sentence structures.
* Use common and basic vocabulary (e.g., use the word “unusual” rather than the word “arcane”).
* Briefly define or explain all technical terms that may be unfamiliar to readers.
* Explain all acronyms the first time they are used in your text—e.g., “Digital Signal Processing (DSP)”.
* Explain local references (e.g., not everyone knows all city names in a particular country).
* Explain “insider” comments. Ensure that your whole audience understands any reference whose meaning you do not describe (e.g., do not assume that everyone has used an Android phone, or a particular application).
* Explain colloquial language and puns. Understanding phrases like “red herring” may require a local knowledge of English. Humor and irony are difficult to translate.
* Use unambiguous forms for culturally localized concepts, such as times, dates, currencies, and numbers (e.g., “1-5- 97” or “5/1/97” may mean 5 January or 1 May, and “seven o’clock” may mean 7:00 am or 19:00). For currencies, indicate equivalences: “Participants were paid ₩22, or roughly US$29.”
* Be careful with the use of gender-specific pronouns (*he*, *she*) and other gendered words (*chairman*, *manpower*, *man-months*). Use inclusive language that is gender-neutral (e.g., *she* *or* *he*, *they*, *s/he*, *chair*, *staff*, *staff-hours*, *person-years*). See the *Guidelines for Bias-Free Writing* for further advice and examples regarding gender and other personal attributes [9]. Be particularly aware of considerations around writing about people with disabilities.
* If possible, use the full (extended) alphabetic character set for names of persons, institutions, and places (e.g., Grønbæk, Lafreniére, Sánchez, Nguyễn, Universität, Weißenbach, Züllighoven, Århus, etc.). These characters are already included in most versions and variants of Times, Helvetica, and Arial fonts.

# Accessibility

The Executive Council of SIGCHI has committed to making SIGCHI conferences more inclusive for researchers, practitioners, and educators with disabilities. As a part of this goal, the all authors are asked to work on improving the accessibility of their submissions. Specifically, we encourage authors to carry out the following five steps:

1. Add alternative text to all figures
2. Mark table headings
3. Generate a tagged PDF
4. Verify the default language
5. Set the tab order to “Use Document Structure”

For more information and links to instructions and resources, please see: <http://chi2016.acm.org/accessibility>.

# Page Numbering, Headers, and Footers

Your final submission should not contain footer or header information at the top or bottom of each page. Specifically, your final submission should not include page numbers. Initial submissions may include page numbers, but these *must* be removed for camera-ready. Page numbers will be added to the PDF when the proceedings are assembled.

# Producing and testing PDF files

We recommend that you produce a PDF version of your submission well before the final deadline. Your PDF file must be ACM DL Compliant. The requirements for an ACM Compliant PDF are available at:

<http://sheridanprinting.com/typedept/ACM-distilling-settings.htm>

When creating your PDF from Word, ensure that you generate a tagged PDF from improved accessibility. This can be done by using the Adobe PDF add-in, also called PDFMaker. Select Acrobat | Preferences from the ribbon and ensure that “Enable Accessibility and Reflow with tagged Adobe PDF” is selected. You can then generate a tagged PDF by selecting “Create PDF” from the Acrobat ribbon. Test your PDF file by viewing or printing it with the same software the publisher will use, Adobe Acrobat Reader Version 10, which is widely available at no cost. Note that most reviewers will use a North American/European version of Acrobat Reader, so please check your PDF accordingly.

# Conclusion

It is important that you write for the SIGCHI audience. Please read previous years’ proceedings to understand the writing style and conventions that successful authors have used. State clearly what you have done, not merely what you plan to do, and explain how your work is different from previously published work, i.e., *the unique contribution that your work makes to the field*. Please consider what the reader will learn from your submission, and how they will find your work useful. If you write with these questions in mind, your work is more likely to be successful, both in being accepted into the conference, and in influencing the work of our field.

# ACKNOWLEDGMENTS

Sample text: We thank all the volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document. Authors 1, 2, and 3 gratefully acknowledge the grant from NSF (#1234-2012-ABC). This is just an example.

# References format

References must be the same font size as other body text. References should be in alphabetical order by last name of first author. Example reference formatting for individual journal articles [3], articles in conference proceedings [7], books [9], theses [10], book chapters [11], an entire journal issue [6], websites [1,4], tweets [1], patents [5], and online videos [8] is given here. This formatting is a slightly edited version of the format automatically generated by the ACM Digital Library (http://dl.acm.org) as “ACM Ref”. More details of reference formatting are available at:

<http://www.acm.org/publications/submissions/latex_style>

Note that the Hyperlink style used throughout this document uses blue links; however, URLs that appear in the references section may appear in black.

# REFERENCES

1. @\_CHINOSAUR. 2014. VENUE IS TOO COLD. #BINGO #CHI2016. Tweet. (1 May, 2014). Retrieved February 2, 2014 from https://twitter.com/\_CHINOSAUR/status/461864317415989248
2. ACM. How to Classify Works Using ACM’s Computing Classification System. 2014. Retrieved August 22, 2014 from [http://www.acm.org/class/how\_to\_use.html](http://www.acm.org/class/how_to_use.html%20)
3. Ronald E. Anderson. 1992. Social impacts of computing: Codes of professional ethics. *Soc Sci Comput Rev* 10, 2: 453-469.
4. Anna Cavender, Shari Trewin, Vicki Hanson. 2014. Accessible Writing Guide. Retrieved August 22, 2014 from <http://www.sigaccess.org/welcome-to-sigaccess/resources/accessible-writing-guide/>
5. Morton L. Heilig. 1962. Sensorama Simulator, U.S. Patent 3,050,870, Filed January 10, 1961, issued August 28, 1962.
6. Jofish Kaye and Paul Dourish. 2014. Special issue on science fiction and ubiquitous computing. *Personal Ubiquitous Comput*. 18, 4 (April 2014), 765-766. <http://dx.doi.org/10.1007/s00779-014-0773-4>
7. Scott R. Klemmer, Michael Thomsen, Ethan Phelps-Goodman, Robert Lee, and James A. Landay. 2002. Where do web sites come from?: capturing and interacting with design history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '02), 1-8. <http://doi.acm.org/10.1145/503376.503378>
8. Psy. 2012. Gangnam Style. Video. (15 July 2012.). Retrieved August 22, 2014 from <https://www.youtube.com/watch?v=9bZkp7q19f0>
9. Marilyn Schwartz. 1995. *Guidelines for Bias-Free Writing.* Indiana University Press.
10. Ivan E. Sutherland. 1963. *Sketchpad, a Man-Machine Graphical Communication System*. Ph.D Dissertation. Massachusetts Institute of Technology, Cambridge, MA.
11. Langdon Winner. 1999. Do artifacts have politics? In *The Social Shaping of Technology* (2nd. ed.), Donald MacKenzie and Judy Wajcman (eds.). Open University Press, Buckingham, UK, 28-40.