Visualizing Competitive Ballroom Results

CS 171 Project Process Book

Robert Krabek and Leila Hofer
April 17, 2015

Table of Contents

1. Overview and Motivation

There are tens of thousands of ballroom dancers around the world who compete regularly in events where they are judged by a panel of professionals against their competitors and given placements accordingly. The need to track ones progress is paramount for dancers aiming to assess the effectiveness of their training regimes or simply to see how they measure up against their peers. The current web-based infrastructure that presents information on dancers' results does not aggregate or display data in a way that makes comparisons over time or between couples easy and intuitive. This project therefore aims to meet the need for an intuitive, user-friendly way for dancers to see where they stand in the world of competitive ballroom.

2. Related Work

This project was heavily inspired by the article that came out in the New York Times highlighting Peyton Manning's record breaking touchdown career. Both the content (showing progress over time) and the style of the visualization (simple line charts with color highlights) seemed appropriate and desirable for our project as we are also aiming to track progress over time and to highlight individuals' progress.

The following is the image that inspired our primary visualization.
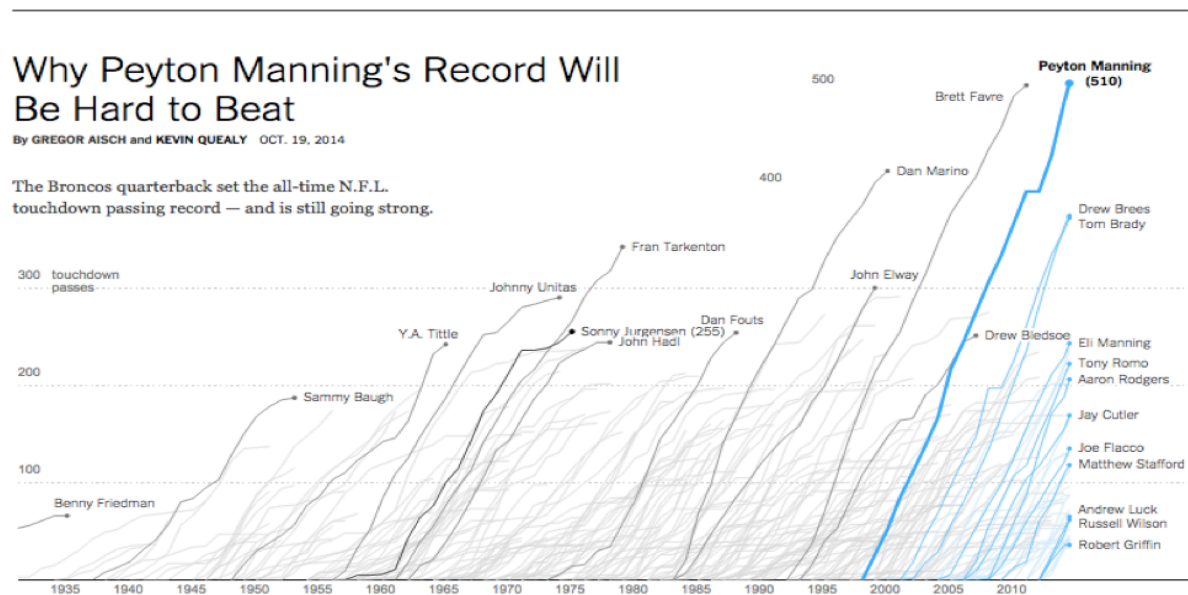
source: http://www.nytimes.com/interactive/2014/10/19/upshot/peyton-manning-breaks-touchdown-passing-record.html?_r=0&abt=0002&abg=1

Instead of touchdowns, our visualization tracks world rankings in ballroom dance based on points compiled through an analysis of results in competition. Instead of individuals, the lines in our dataset track couples. Our visualization has the option of letting the user customize their visualization to their needs. There is functionality for the user to opt to display data that compares them to specific couples of thier choosing and to compare them

to all of the dancers in the dataset. The simplicity of this design combined with its ability to convey a large amount of information effectively is why we were attracted to it. We use bolded lines and color when necessary to highlight couples in question.

3. Driving Questions

Summary
- How has my score as a couple changed throughout my dancing career?
- How has my score changed compared to other couples I know?
- How have specific competitions affected my score as a dancer?

Questions Expanded
As stated above, this project aims to allow dancers to track their progress and compare themselves to other dancers over time. These comparisons will be customizable so that dancers can either compare themselves to other couples. To contextualize these comparisons over time, we ground users of our product in the results of specific competitions so that they can see how others did at specific points relative to themselves. This will be accomplished first by displaying results from competitions at a user's request. Second, so that dancers know how their local field compares to the ballroom community as a whole, we intend to integrate a feature to allow couples to view all of the couples in the dataset.

Our goal is to produce a visualization that is customizable and answer a variety of questions about dancers' progress over time and relative to others. To accomplish this we intend to implement a set of coordinated displays pertaining both to the individual dancer and to overall trends in the ballroom world. Through this project we hope to better our understanding of how to manipulate and present data with JavaScript and d3 and to create a useful product for ballroom dancers around the world.

4. Data

About
The primary data we are interested in comes from DancesportInfo.net, one of the only definitive sources of data for competitive ballroom results. DancesportInfo's databases currently contain results from 26,353 competitions in 91 countries for over 20,641 registered dancers. The API for procuring data from DancesportInfo.net is linked below. The API allows extraction of the necessary data for this project, e.g. competitor names, results, competitions, etc. The credit for the API work goes to a friend and ballroom dancer Cloud Cray (GitHub Gist page link here).

Data Processing
There was a significant amount of aggregating and reshaping that was necessary before we could begin visualizing our data. The data we procured was from 70 competitions, each one contained in a different json file. The folder /dancesportinfo_results contains the 70 files along with a python script, script.py, that concatenates all 70 files and outputs a single json

file. Afterwards we needed to reshape the data from being organized by competition to being organized by couple in order to make it useful for our primary. The reshaping process takes data with the form:

```
[
        {
                competition name:
                competition id:
                competition events: [
                        {
                                event style:
                                event date:
                                event results: [
                                        {
                                                result:
                                                couple:
                                                couple score:
                                        },
                                        …
                                ]
                        },
                        …
                ]
        },
        …
]
```

and gives it the structure below.

```
[
        {
                couple name:
                couple id:
                couple events: [
                        {
                                competition name:
                                competition id:
                                event style:
                                couple score:
                        },
                        …
                ]
        },
        …
]
```

We chose this structure because it allows us to index the data by couple and have events objects ready for each couple so that if that couple is chosen for visualization we have ready an object for each competitor with all of their scores and the date that they received the scores. When we began the visualization process we found that we had over 1500 unique couples from over 70 competitions. Additionally, data had to be sorted on several levels while reshaping.

All of this is done in index.html in the /data_and_implementations folder as javascript does not handle client-side work very easily in order to produce a separate json file.

5. Exploratory Data Analysis

We originally experimented with the C3 charting library that is based on D3 but provides more streamlined, simplified code. Our first prototype of visualization 1 was generated in C3 for two couples at different competitions. Although it provided an easy means of charting the data, C3 did not prove versatile enough to implement the fully integrated designed view that we were aiming to produce so we began our second prototype using the D3 library.

6. Design Evolution

Visualization 1: Main View
Or first visualization was initially intended to show a large number of couples at once as the NY Times image shown above suggests. However, we realized that unlike their data which showed rapid growth over a short time period, our data was from a much shorter time period and so displayed much more horizontal trends. This means that individual couples' lines overlap quite a bit when all the data is shown, making this not the most useful display. We therefore chose to set our default load screen to a single couple in order to avoid clutter and allow individuals to focus on individual couples

Visualization 2: Supporting View
Our second integrated view was originally intended to be an overlapping bar chart showing total score before and after but upon experimentation we decided that the change was clearer if appended to the end of a single bar, making a stacked bar chart. This also allowed us to use color highlights to accent changes in intuitive ways (e.g. red for decreased score, green for increased score, yellow for no change).

7. Implementation

A note about the files
- Folder /dancesportinfo_results contains:
  - The raw data in 70 json files
  - script.py, a python script for concatenating the raw data into a single file
- Folder /data_and_implementations contains:
  - Folder /data

- combined.json, json object holding the concatenated raw data
        - index.html, the home page of our visualization
        - Folder /js
            - scorevis.js, our primary visualization
            - resultvis.js our integrated view visualization
        - Folder /libs
            - Libraries for the project
    - Folder /Images contains:
        - Images for the .md files
    - Process_book_draft.md for milestone 1

Other files are largely extraneous.

Visualization 1: Main View
- On load this view displays:
    - Scores over time for the first couple in our dataset in a line chart
    - Points representing events at which the couple has danced and been scored
    - A dropdown menu with this couple's name selected and three buttons with
    - Three buttons with functionality to add another dropdown menu to select another couple, delete the last selection (if no more couples remain the visualization will display all couples), or to display all couples directly
        - Displaying all couples takes approximately 4-5 seconds to load
- On hovering over the line:
    - The line is highlighted
    - The couple's name appears at the end of the line
- On hovering over a point:
    - A tool tip reports what competition the point is from
- On clicking a point our second visualization is triggered

Visualization 2: Supporting View
- On clicking a point on visualization 1 the placements from the event that the point represents are displayed in a bar chart
- On hovering over a bar:
    - Colors of the names displayed in the bar and colors of bars themselves change slightly to highlight the couple
    - Translucent negative changes at the end are made opaque
    - A tool tip displays the change in points a couple experienced as a result of their placement and the couple's new score
    - Colors of tool tip are different depending on whether a couple gained points, lost points, or kept the same number of points.

8. Evaluation

Future Improvements

Overall the visualization works well without any major glitches. Our one minor concern is the time it takes to load originally (approximately 3-4 seconds) and the time it takes to go from a display of a few couples to showing all couples (approximately 4-6 seconds).

To improve the scope of our visualization and make it more useful for dancers, we would have loved to incorporate data from a longer time horizon and from more countries as most of our data comes from the US in the last twelve months. More data, over a longer period of time would provide a fuller understanding of trends over time. Future versions of this project might also include functionality to compare a specific competition for a couple to their performance on average in competition, in terms of how many points they gain or lose.

Unexpected realizations
In the process of building and using our visualization we realized that progress in ballroom is much more stochastic than we originally hypothesized. In the long run scores do tend to increases over time, but in between there are large and unexpected jumps. Overall, much to our surprise, our data's trends resemble the stock market much more than they resemble the NY Times touchdown data that we originally based our visualization on.

9. References

D3: http://d3js.org/

Jquery: https://jquery.com/

D3.queue: http://giscollective.org/d3-queue-js/

Bootstrap Library: http://getbootstrap.com/javascript/

D3-tip: https://github.com/Caged/d3-tip

Reference to C3: http://c3js.org/

Google Fonts: https://www.google.com/fonts

Font Awesome Icons: https://fortawesome.github.io/Font-Awesome/icons/