Assignment 4: Becoming an Independent Data Scientist

1 Region and Domain

State the region and the domain category that your data sets are about.

**Wheaton, Illinois, United States**

**Sports and Athletics**

2 Research Question

You must state a question about the domain category and region that you identified as being interesting.

**How have the win percentages for the five major Chicago sports teams (The Chicago Bulls, The Chicago Bears, The Chicago Blackhawks, The Chicago Cubs, and The Chicago White Sox) change since 1979?**

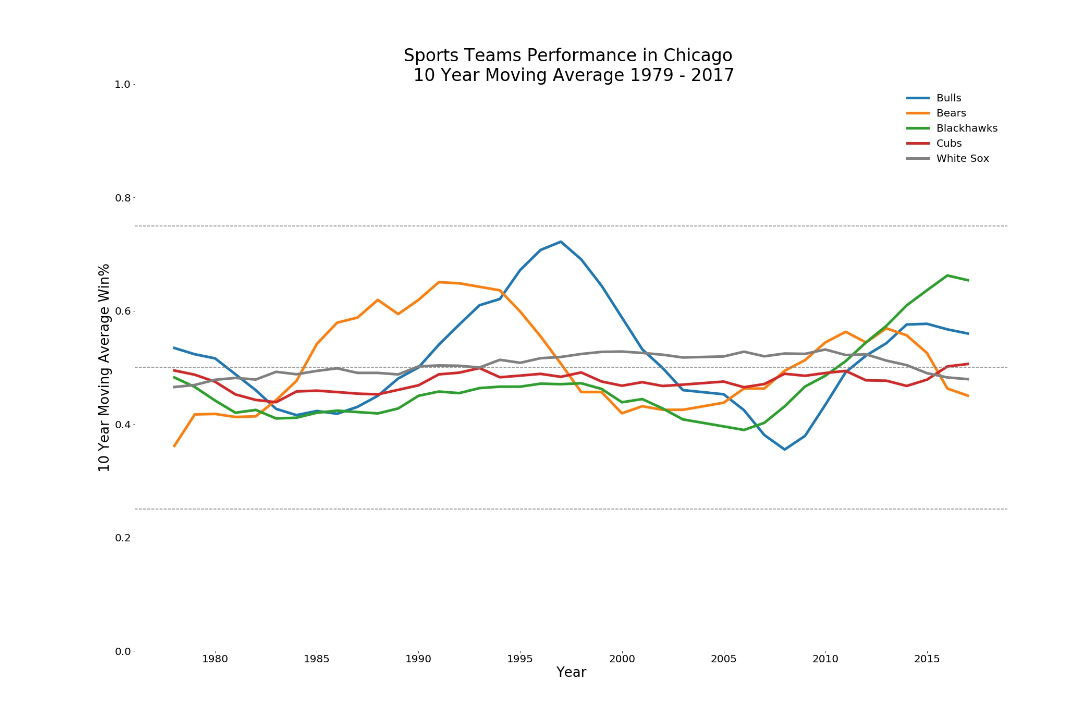
3 Links

You must provide at least two links to publicly accessible datasets. These could be links to files such as CSV or Excel \_les, or links to websites which might have data in tabular form, such as Wikipedia pages.

* <https://en.wikipedia.org/wiki/List_of_Chicago_Bulls_seasons>
* <https://en.wikipedia.org/wiki/List_of_Chicago_Bears_seasons>
* <https://en.wikipedia.org/wiki/List_of_Chicago_Blackhawks_seasons>
* <https://en.wikipedia.org/wiki/List_of_Chicago_Cubs_seasons>
* <https://en.wikipedia.org/wiki/List_of_Chicago_White_Sox_seasons>

4 Image

You must upload an image which addresses the research question you stated. In addition to addressing the question, this visual should follow Cairo's principles of truthfulness, functionality, beauty, and insightfulness.



5 Discussion

You must contribute a short (1-2 paragraph) written justi\_cation of how your visualization

addresses your stated research question.

This visualization was focused on answering the question of How have the win percentages for the five major Chicago sports teams (The Chicago Bulls, The Chicago Bears, The Chicago Blackhawks, The Chicago Cubs, and The Chicago White Sox) change since 1979. I scraped for data from Wikipedia about the teams wins and losses by season for each team. Unlike the example, I did not drop tie games of the analysis as I think the resulting number is more accurate. Moving averages were provided to display teams’ trends rather than specific outcomes.

The plot indicates The Chicago Bulls was the team with the highest win percentages among the teams. Chicago Bulls was the highest performer in the group from 1994 to 2002 – moment the team underperform the rest of the group. The Chicago Bears is the second leading win percentages during the period 1985 to 1994. Both teams experienced a downward trend after 1998. The Chicago Blackhawks is the current top performer in the group, due to its upward trend since early 2000’s.

* Describe your design choices for your visual in regards to Cairo's principle of **truthfulness**.

The data used in my visualization was neither cherry-picked nor changed to fit designer’s assumptions. The method used to calculate the number across the series was expressly mentioned and applied in a similar fashion on each team. Also, there is no truncation in both y and x axes.

* Describe your design choices for your visual in regards to Cairo's principle of **beauty**.

The graph considered that a sign of beauty is a decluttered graph. Spines and grids were taken out from the graph to help the audience see the team’s performance in question

* Describe your design choices for your visual in regards to Cairo's principle of **functionality**.

Functionality has been provided by decluttering the graph and properly labeling each team. I included dash lines to help the audience easily both understand the range of success of the selected teams, and noticed the periods in which teams were above 50% Win%. Finally color was used to differentiate among teams.

* Describe your design choices for your visual in regards to Cairo's principle of **insightfulness**.

By drawing dash lines and clearly differentiating each team, I facilitate the audience to reach an Eureka moment. Both, outperformers and underperformers can be easily noticed. Thresholds also contribute differentiate between not only outperformers and underperformers, but also differenciate outliers.