Ibrahim Hassanain,Tszchung Cheung, Amrit Singh,

Hamzah Saleh, Lifu Tao, Shazid Rahman

City College of New York

Professor Retemadpour

31 March 2021

Group Assignment 1

<div align="center">Group Assignment Surprise Map</div>

Below we have added links to all the files via web in case the zip folder contents are not working
as intended.

Link to GitHub (if needed): https://github.com/asingh2066/Visualization-Group-Assignment1.git

Link to Graph: https://colab.research.google.com/drive/1mEQ3qB8ewh_xOq5Xltl-CI0txoS24UXL?usp=sharing

Link to Video: https://www.youtube.com/watch?v=3nr0Bj-Hn1M

      The 2019-2020 NBA season was the last season we would ever witness before the
pandemic came into play. Under no unusual circumstances, players and teams were competing at
a high level for the NBA championship. Intuitively, the more points a team averages, the more
likely they will win. We have created two Choropleth maps below to show the teams' average
efficiency during last season (no Covid-19) and this season (with Covid-19). Efficiency is
expressed by a stat referred to as 'efficiency' and abbreviated EFF. It is derived by a simple
formula: `(PTS (points) + REB (rebounds) + AST (assists) + STL (steals) + BLK (blocks) — Missed FG (field goals) — Missed FT (free throws) — TO (turnovers) / GP (games played)`. There are 29 teams in the United States and 1 team
in Canada. This visualization captures all 29 teams in the United States based off their respective
locations: states. A few states might have more than one team and if so, the average efficiency of
all teams in one state will be accounted for to come up with one average for the whole state
entirely.

      When it came to showcasing some relative values, it was a bit tougher for a choropleth
map given our data and the structure. So instead, we decided to organize the mean/average
relative values for the columns based on overall performance of the states and the respective

teams in each state. The relative values being shown on the choropleth map now would be the average overall efficiency of each state and their teams. Looking at the graph you can see that efficiency is color coded for each state, lighter shaded meaning less efficient compared to the darker shades of red. Hovering over the states you can see more data on other mean relative values we calculated within the given data, and overall, this choropleth shades the states on the map based on our area of interest within the dataset, here being Efficiency.

As you can see below, the first map shows the 2019-2020 season. These averages are mainly expected since there was no pandemic during these times. The darker a state gets, the more "hotter" or the more efficient that state has averaged based off its team(s). Also, if you hover over the states, you will see the Field Goal percentage that state may average as well. There is no direct relationship between average efficiency and average FG%. Usually a higher FG% will indicate higher efficiency and in most cases, it can correlate to the fact that the state may have a high average of points as well. Based on the 2019-2020 season, the state of Oregon had the highest averaged efficiency while the state of Oklahoma had the least averaged efficiency that season. It is also to note that Oregon did not have the most points in the whole season although their averaged efficiency was the highest.

Focusing now on the Surprise portion of the map, sports data was not the best to illustrate regional data, so we had to go with state data instead. For the choropleth portion we had mean relative values, however these values also provide statistics that you would not expect for some teams, coming as a surprise. Before getting into that, we tried to implement the Formal Bayesian Theory when it came to analyzing the surprise portions for the data. However, after trying for a while, it was a bit difficult with our dataset, so we decided not to do any advanced mathematical calculations. Instead, we filtered our data to get mean values that we needed and base our conclusion on that. Some things that came as a surprise were how certain state teams have the highest efficiency rating but struggle to make it to final games and win championships, and how that various teams have increased or have consistent performances during the covid period. An example of this would be during the 2019-2020 season, Oregon had the highest overall efficiency, however, California won the 2020 finals.

Looking at the 2020-2021 season map, you will start to see that there have been some significant changes to the anticipated values that we usually get during regular seasons. These values are not only impacted by covid but have altered due to many factors such as injuries and

player trades, which can be seen with states such as California, Arizona, Wisconsin, and Illinois. However, some other states are doing better despite having covid around such as Texas, Florida, Oklahoma, and Virginia, while some other states are doing worse because they are heavily affected by Covid, for instance Oklahoma. Something that stands out as an outlier in this dataset would be how the number of overall points does not conclude whether your team is in the top 5 standings, and an example of this can be seen with Washington, Oregon where the team is averaging around 9000 points but are not winning championships. Some other data that may not be expected would be performance of teams with top players, such as the Brooklyn Nets, NY. Looking at the covid map, New York was underperforming in points and efficiency despite having multiple top players on the roster. Usually, it is expected for teams to overperform and dominate in their region after acquiring such players, but in this data, it is surprising to see this kind of performance after looking at the statistics.