Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**

EIT Data Science - Master in Digital Innovation

Data Visualization

Group 9 - Final Assignment
Using Data Visualization for MMA Fight Analysis

Máté Lukács
Ádám Földvári
Inés Simón del Collado
Davis Siemens

Madrid, 12 January 2025

Contents

1. Problem Characterisation

The Ultimate Fighting Championship (UFC) is the largest global promoter of fighting sports. They specialized in MMA, a combat sport where athletes use techniques from disciplines such as wrestling, jiu-jitsu, boxing, and kickboxing. As a relatively young and rapidly evolving sport, the UFC has seen great shifts in its organization over the last decades. This rapid evolution coupled with the complexity of the sport itself, raises several important questions for analysts, fighters, and coaches alike.

Within this project, we have crystallized four of those and created corresponding idioms.

- Idiom 1: **Trend Analysis of Fight Finishes**
- Idiom 2: **Exploring Fight Outcome Distributions Across UFC Weight Classes**
- Idiom 3: **UFC Fight Performance Heatmap**
- Idiom 4: **Geographical Distribution of UFC Fights**

1.1 Dataset Description

For the visualization, we are using the Ultimate UFC Dataset[1]. It is a combination of multiple UFC datasets. Includes every UFC fight from mid 2010 until the present. It is a structured dataset with rows refer to individual bouts and columns offer features of the fighters and the fight itself.

The dataset represents a flat map and is the most complete UFC dataset currently publicly available. It contains 4,838 entries and 119 attributes, with an overall high data quality. The dataset covers diverse aspects of UFC events, including fighter statistics, fight outcomes, and event details. In the following chapter, we elaborate which data attributes have been used in the idioms creation.

R_fighter	B_fighter	date	location	country	Winner	title_bout	weight_class
Jared Cannonier	Kelvin Gastelum	8/21/2021	Las Vegas, Nevada, USA	USA	Red	FALSE	Middleweight
Clay Guida	Mark Madsen	8/21/2021	Las Vegas, Nevada, USA	USA	Blue	FALSE	Lightweight
Parker Porter	Chase Sherman	8/21/2021	Las Vegas, Nevada, USA	USA	Red	FALSE	Heavyweight
Trevin Jones	Saidyokub Kakhramonov	8/21/2021	Las Vegas, Nevada, USA	USA	Blue	FALSE	Bantamweight
Vinc Pichel	Austin Hubbard	8/21/2021	Las Vegas, Nevada, USA	USA	Red	FALSE	Lightweight
Alexandre Pantoja	Brandon Royval	8/21/2021	Las Vegas, Nevada, USA	USA	Red	FALSE	Flyweight
Austin Lingo	Luis Saldana	8/21/2021	Las Vegas, Nevada, USA	USA	Red	FALSE	Featherweight
Brian Kelleher	Domingo Pilarte	8/21/2021	Las Vegas, Nevada, USA	USA	Red	FALSE	Bantamweight
Bea Malecki	Josiane Nunes	8/21/2021	Las Vegas, Nevada, USA	USA	Blue	FALSE	Women's Bantamweight

Figure 1.1: Exemplary Snipped of Ultimate UFC Dataset.[1]

2. Idiom 1 - Trend Analysis of Fight Finishes

In the UFC fighters can finish a match using a variety of techniques (e.g., submissions such as chokes, or strikes such as punches or kicks). Because the sport is still maturing, it is important to question how frequently certain finish types occur over time and how those finishing methods proportionally compare in a given year to help stay on top of the sport.

2.1 Task Definition

- **Task**: Analyse the evolution of finish types in UFC fights over time, focusing on which finishing categories (Chokes, Kicks, Strikes, Joint Locks, etc.) are used most often, and what fraction of the total finishes each category represents in a given year.
- **Purpose**: We want to discover trends and patterns (e.g., "Are choke finished increasing or decreasing in frequency compared to other finishes?" "Are finishes by kicks more prominent in modern years compared to non-striking finishes?"). Especially important here is, to understand if an increase in one finish type, comes with a decrease of an other one. This knowledge can be used by analyst, fighters or coaches.

This overall problem characterization frames the basis for our visual analytics tool: the user needs to explore the dataset, filter by year range, compare raw counts vs. relative proportions, and potentially smooth out short-term fluctuations to see long-term trends.

2.2 Data Abstraction

The used dataset is elaborated on in Section 1.1. For the described task we will use the following columns:

- *Index*: Ordinal key attribute.
- *Date*: Temporal attribute, used to retrieve the year of the fight.
- *RedFighter*: Categorical attribute (String) representing name of the first fighter
- *BlueFighter*: Categorical attribute (String) representing name of the second fighter
- *FinishDetails*: Categorical attribute describing the finishing move (e.g., *Punch*, *Rear Naked Choke*). *NaN* if fight was not finished early.

- *Category* (derived): Hierachical grouping of *FinishDetail* in *Chokes*, *Strikes*, *Kicks*, *Joint Locks* or *Other*.
- *FightID* (derived): Value to identify fights, derived by *RedFighter*, *BlueFighter* and *Date*. The FightID is used as a key in the algorithmic implementation.

Hence, we have table data with categorical (finish category) and temporal (year) dimensions. The main measure is the count the proportion of a *Category* or *FinishDetail* relative to the total finishes that year. Second measure is the total count of a specific *Category* or *FinishDetail* for a respective year. The below table shows the derivation of the *Category* by *FinishDetail*.

Finish Type Category	**FinishDetail(s)**
Chokes	`Guillotine Choke, Rear Naked Choke, Anaconda Choke, North-South Choke, D'Arce Choke, Triangle Choke, Arm Triangle, Peruvian Necktie, Scarf Hold, Von Flue Choke, Ezekiel Choke, Inverted Triangle, Other - Choke`
Joint Locks	`Armbar, Straight Armbar, Triangle Armbar, Kimura, Heel Hook, Kneebar, Ankle Lock, Omoplata, Keylock, Other - Lock`
Kicks	`"Kick" + "Kicks" → "Kick(s)", "Knee" + "Knees" → "Knee(s)", Flying Knee, Spinning Back Kick`
Strikes	`"Punch" + "Punches" → "Punch(es)", "Elbow" + "Elbows" → "Elbow(s)", "Spinning Back Elbow" + "Spinning Back Fist" → "Spinning Back Strike"`
Other	`Slam, Takedown, Other`

Table 2.1: Listing of all *Categories* and *FinishDetails*. Some *FinishDetails* have been merged as these are very similar in the targeted domain (fighting sport).

2.3 Task Abstraction

This project focuses on the target of **trend analysis** and **comparison** over time of finishing techniques in the UFC. The core questions revolve around understanding whether certain finish types—such as chokes, strikes, kicks, or joint locks—are gaining or losing popularity from year to year. Specifically, we want to see the relative proportions of these finish categories as a fraction of total finishes for each year and to compare their absolute counts if desired. By doing so, they can determine, for example, whether chokes are more frequent in modern MMA, whether certain striking finishes are less common than they once were, or whether joint locks have seen a resurgence over time. These inquiries reflect a **knowledge-generation** action (**discover**), as the user aims to discover new insights (e.g., long-term trends, surges, or declines) that cannot be easily captured by simpler, non-visual queries.

2.4 Visual Encoding

We use a stacked area chart to arrange data over time. The x-axis represents *Year*, and the y-axis encodes *proportions* (stacked to 100% per year) or alternatively the *raw count* of each *Category* or *FinishDetail*. Each of the category or FinishDetail

is visualized as a distinct area segment, arranged cumulatively. This layout allows users to assess both the overall total finishes and each category's contribution.

The arrangement of data follows best practices for temporal trends, with the time dimension naturally progressing from oldest (left) to newest (right). The stacking ensures a clear part-to-whole relationship over time, adhering to Munzner's Expressiveness principle for data representation. [2]

We encode:

- Visual Marks: Each category is represented by a colored area segment (two-dimensional). The height of the area encodes either the raw count of finishes or its relative proportion. Additionally, the line (one-dimensional) of the time series reflects increases or decreases over time.

- Color Channel: Categories are assigned distinct hues to differentiate them, ensuring immediate recognition. A categorical color scale was choosen.

- Vertical Position: The height of the stacked areas encodes either the absolute value (raw count) or normalized proportion (percentage) of finishes for each category.

- Horizontal Position: The horizontal axis encapsulates the year information.

This mapping adheres to Shneiderman's guidelines for clear labeling and contrasts, enabling users to focus on the data attributes. [3]

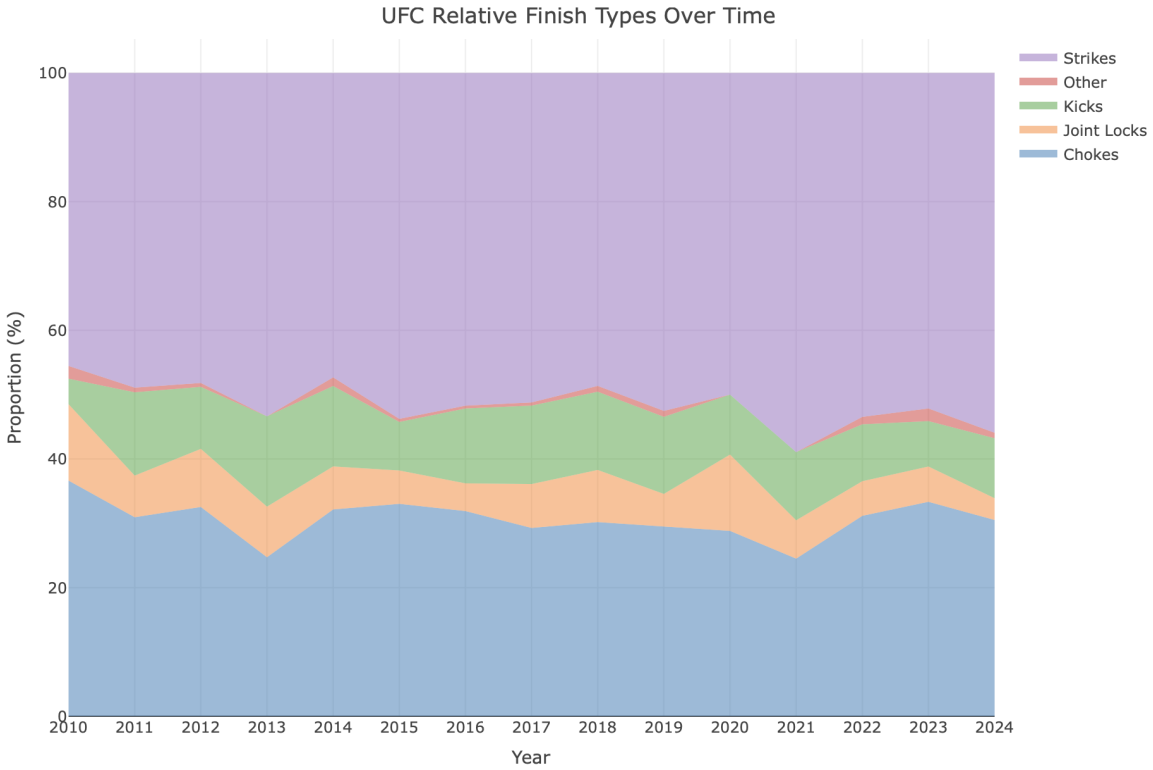

Figure 2.1: Default view of Idiom 1.

2.5 Interactions

2.5.1 Manipulations

2.5.1.1 Raw Count Toggle

While the default view in Figure 2.1 shows the proportions of *Categories* or *FinishDetails* the user can decide to switch to a raw (total) count. This then displays absolute numbers of finishes for each category per year. This toggle allows the user to shift focus between **absolute trends** and **relative comparisons**.

2.5.1.2 Hover Text-box

Users can hover over the stacked areas to view precise details for a specific year and category. Tooltips display both the raw count and the proportion, supporting “detail on demand” as described by Shneiderman.[3]

2.5.1.3 Year Range Slider

The visualization includes a slider to set the minimum and maximum years. This lets users zoom in on specific time periods.

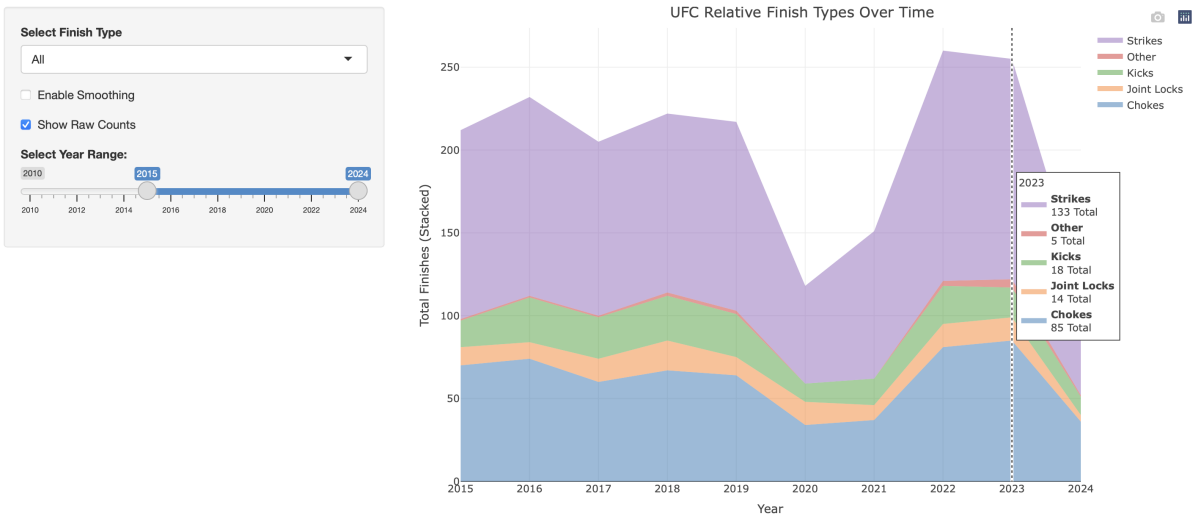

Figure 2.2: Idiom 1, after activating "Show Raw Counts" Toggle and selecting a specific time range. Additionally, the hover text-box is shown to give detailed information for a selected year.

2.5.1.4 Statistical Smoothening/ Data Analysis Algorithm

In our visualization, we incorporate LOESS (Locally Estimated Scatterplot Smoothing) to reveal better reveal long-term trends. Doing so we have included an analysis algorithm.

This smoothing option, controlled by a user toggle ("Enable Smoothing"), allows users to focus on broader patterns and trends over time, helping to distinguish between real long-term shifts and random year-to-year variations, without disrupting data patterns. An example can be seen in Fugure 2.3.

Figure 2.3: Idiom 1, comparing smoothended and unsmoothened visualization while filtering for category "*Joint Locks*". Through the smoothening we are able to recognize a trend of "Other" joint-locks getting more popular recently. This underlines the idea that new submission techniques are getting developed and used in the octagon in the recent years.

2.5.2 Reductions

2.5.2.1 Category Deep Dive

The visualization includes an interactive category filtering feature, allowing users to focus on a specific finishing category (e.g., “Kicks” or “Chokes”) by selecting a drop down list. This functionality enables users to explore detailed trends within a particular category while reducing visual complexity.

When a specific category is selected, the tool automatically extracts its subcategories (specific *FinishDetails*) and displays only the top 5 most frequent *FinishDetails* for that category.

This can be shown in an example: If the user selects “*Kicks*” in the drop down, the most frequent techniques (e.g., “*Kick(s)*,” “*Knee(s)*,” “*Flying Knee*”) will appear, sorted by their contribution to the total number of finishes. Any additional subcategories beyond the top 5 are aggregated into an “Other - Kicks” group to keep the chart lean.

This approach aligns with visualization best practices by prioritizing the most relevant information and avoiding visual clutter.

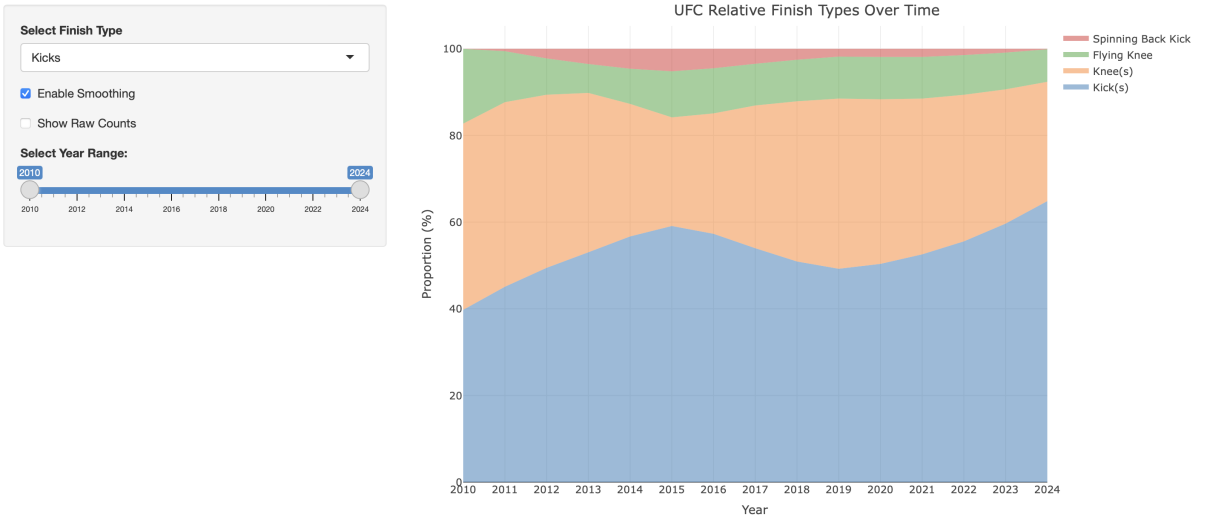

Figure 2.4: Idiom 1, filtering for *Category* "Kicks".

2.5.2.2 Trail Isolation

By clicking on a *Category* or *FinishDetail* in the legend of the graph, the corresponding trail can be isolated. Doing so, the user can deselect, e.g., dominant *FinishDetails* to better focus on the relationships of non-dominant *FinishDetails*.

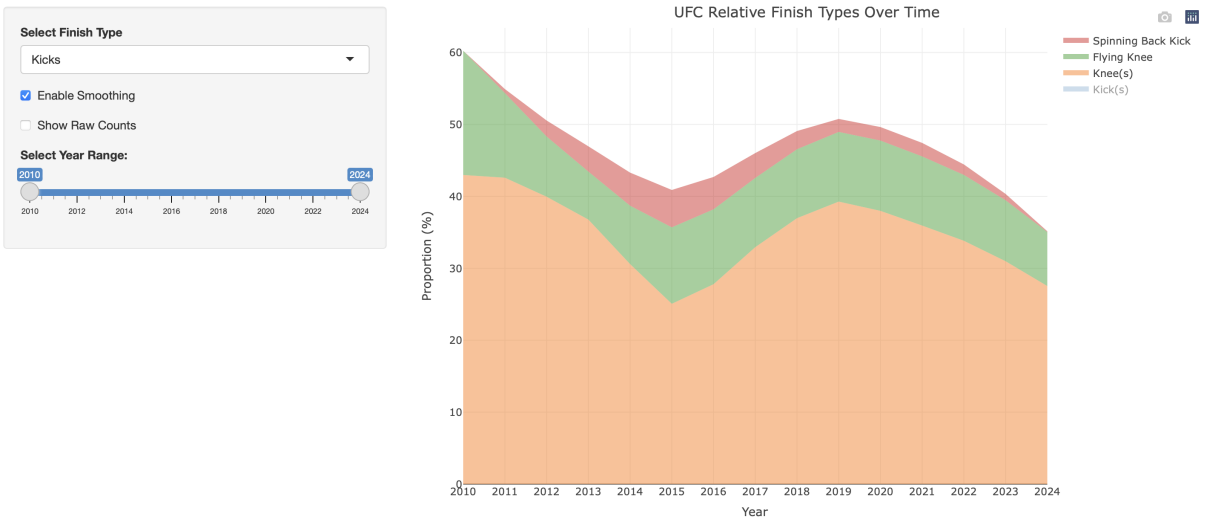

Figure 2.5: Idiom 1, filtering for *Category* "Kicks" and isolating *FinishDetail* "Kick(s)".

2.6 Algorithmic Implementation

The implementation of Idiom 1 can be summarized in four components.

2.6.1 Data Cleaning

The input data, containing details of UFC fights, was filtered to exclude any records with missing or invalid values in the `FinishDetails` column. The `Date` column was converted into a standardized `Date` format and used to extract the `year` as a numeric attribute, enabling temporal analysis. To avoid counting fights multiple

times, a unique identifier (`FightID`) was created by grouping records based on attributes such as `Date`, `RedFighter`, and `BlueFighter`. Additionally, singular and plural finishing details (e.g., `"Punch"` and `"Punches"`) were standardized into unified labels (e.g., `"Punch(es)"`), and related finishing techniques were categorized into broader groups (e.g., `"Rear Naked Choke"`, `"Triangle Choke"` → `"Chokes"`). Less frequent subcategories were grouped into an `"Other"` category when necessary to reduce visual clutter.

2.6.2 Data Aggregation

The algorithmic implementation of the visualization begins with summarizing the data by grouping it by year and category while calculating the count as `n_distinct(FightID)` to ensure each fight is counted only once per finishing category, avoiding redundancies. For each year, the total finishes `(year_total)` are computed, and the proportion for each category is derived, enabling the visualization of both raw counts and relative proportions.

2.6.3 LOESS Smoothing

When the show_smoothing toggle is activated, LOESS smoothing `(span = 0.75)` is applied to reduce short-term fluctuations and reveal long-term trends. For proportions, the smoothed values are re-normalized to ensure they sum to 100% within each year, maintaining the integrity of the part-to-whole visualization, while raw counts are clipped at zero to avoid nonsensical values.

2.6.4 Stacked Plot

The stacked area chart is rendered using `plotly`. A `hovertemplate` is implemented to dynamically display both the percentage and raw total for each category, providing precise details on demand.

Native *shiny* and *plotly* features allow to implement select category filters, dynamically switching between “All” categories or a specific category with its top 5 subcategories (aggregating the rest into “Other”), a year range slider, etc.

3. Idiom 2 - Exploring Fight Outcome Distributions Across UFC Weight Classes

3.1 Task Definition

The primary goal of this idiom is to explore how fight outcomes are distributed across different weight classes in the UFC. Analysts can investigate the prevalence of various finish types (e.g., KO/TKO, Submission, Unanimous Decision, etc.) and how these outcomes vary based on gender, title bouts, and specific user-selected finish types. The interactive nature of the visualization allows users to filter and refine their focus, answering key questions such as:

- Which weight classes have the highest number of KO/TKOs or submissions?
- Do title bouts exhibit different outcome distributions compared to non-title fights?
- Are there differences in fight outcomes between male and female fighters?

This idiom supports exploratory analysis, enabling analysts to discover patterns and trends in fight outcomes across weight divisions.

3.2 Data Abstraction

The dataset used includes attributes relevant to fight outcomes:

- *WeightClass*: Categorical attribute representing the division (e.g., Bantamweight, Lightweight).
- *Finish*: Categorical attribute describing the fight outcome (e.g., KO/TKO, Submission).
- *Gender*: Categorical attribute indicating whether the fighters are male or female.
- *TitleBout*: Boolean attribute specifying whether the fight was a title bout.

The data is aggregated by weight class and finish type, with counts representing the number of fights for each combination. Filters for gender, title bouts, and specific finish types allow users to refine their analysis.

3.3 Task Abstraction

This idiom is aligned with the actions of **discover** and **compare**.

In one way by **discovering** users can identify which weight classes favour specific finish types or detect patterns in gender-based or title-bout-specific outcomes.

On the other hand, by **comparing** analysts can recognize the prevalence of different finish types across weight classes or between male and female fighters.

3.4 Visual Encoding

The visualization employs a **radar chart** to encode the data:

- **Axes:** Each axis represents a UFC weight class (e.g., Featherweight, Middleweight).
- **Lines/Polygons:** Each finish type (e.g., KO/TKO, Submission) is represented by a distinct line or polygon on the radar chart.
- **Color:** Finish types are encoded using distinct colors (e.g., red for KO/TKO, blue for Submission) to ensure clear differentiation.
- **Position:** The distance from the center along an axis represents the count of fights with a specific finish type in that weight class.

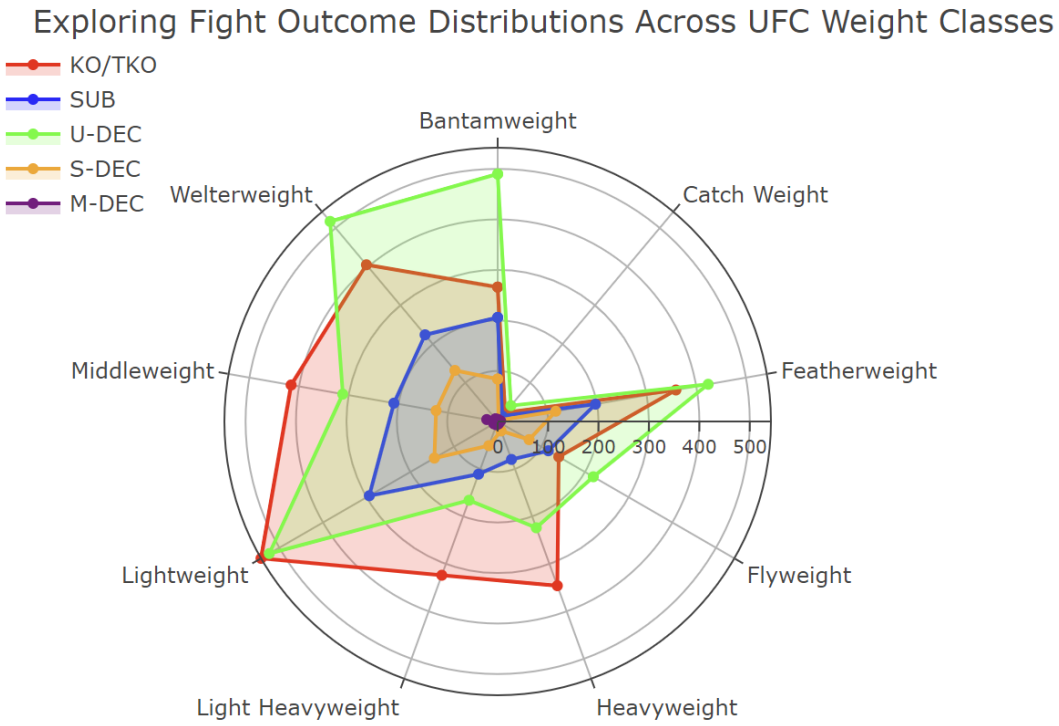

Figure 3.1: Default view of Idiom 2.

This encoding allows users to quickly assess which finish types dominate specific weight classes and how these distributions vary across filters.

3.5 Interactions

The radar chart visualization provides a range of interactive features that enhance user engagement and analytical depth. These interactions allow users to explore fight outcome distributions across weight classes in a highly customizable and intuitive manner. Below are the key interactive elements:

3.5.1 Gender Filter

Description: A radio button allows users to filter the data based on the gender of the fighters: Male or Female.

Purpose: This filter enables analysts to compare fight outcome distributions between male and female fighters, uncovering potential differences in trends or patterns.

Implementation: The 'input$gender' value dynamically filters the dataset, ensuring only fights matching the selected gender are included in the radar chart.

3.5.2 Title Bout Filter

Description: A toggle switch labeled *"Show Title Bouts Only"* lets users focus exclusively on title fights or include all bouts.

Purpose: This feature helps analysts assess whether title bouts exhibit different outcome distributions compared to non-title fights.

Implementation: When enabled, the 'input$boutTypeSwitch' restricts the dataset to fights where `'TitleBout == TRUE'`.

3.5.3 Finish Type Selection

Description: A checkbox group allows users to select specific finish types (e.g., KO/TKO, Submission, Unanimous Decision). Two buttons (*Select All* and *Deselect All*) provide quick options for managing selections.

Purpose: This functionality helps users focus on specific finish types of interest, enabling detailed analysis of their distribution across weight classes.

Implementation:

- The `'input$finishTypes'` determines which finish types are displayed on the radar chart.
- The *Select All* and *Deselect All* buttons use observers (`'observeEvent'`) to update the selected checkboxes programmatically.

3.5.4 Legend Interaction

Description: Users can click on items in the legend to isolate or hide specific finish types (basically this is similar to 3.5.3 Finish Type Selection, just another way to filter the chart).

Purpose: This feature reduces visual clutter and allows users to focus on a single finish type or compare a subset of finish types without distraction.

Implementation: The legend functionality is handled natively by Plotly, which updates the visualization dynamically when legend items are clicked.

3.5.5 Hover Tooltips

Description: When users hover over points on the radar chart, a tooltip appears displaying detailed information about that point. The tooltip includes:

- The weight class (e.g., Bantamweight, Lightweight).
- The finish type (e.g., KO/TKO, Submission).
- The count of fights for that finish type in the given weight class.

Purpose: This "detail-on-demand" interaction provides precise values for each data point, supporting deeper analysis without overwhelming the user with excessive labels.

Implementation: Tooltips are implemented using Plotly's `'hovertemplate'`, which dynamically generates content based on the hovered point's data.

3.5.6 Zoomable Axes Range

Description:

- Users can zoom in or out on the radial axis by adjusting its range dynamically.
- This feature allows them to set custom minimum and maximum values for better focus on specific data ranges.

Purpose:

- Helps users analyze data at different levels of granularity by focusing on smaller ranges (e.g., low fight counts) or expanding to see broader trends (e.g., high fight counts).
- Reduces visual clutter by scaling the chart appropriately for datasets with varying distributions.

Implementation: The radial axis range is configured dynamically using Plotly's `layout.polar.radialaxis.range` property, which adjusts based on user input or zoom actions.

3.6 Example Use Cases

The **first image** (Figure 3.2) represents the default setup of the filters, where the user has selected Male fighters, enabled the Show Title Bouts Only toggle, and chosen finish types `'KO/TKO','Submission','Unanimous Decision'`. This configuration provides a filtered view of the radar chart, displaying fight outcome distributions across all UFC weight classes without further adjustments to the axis range.

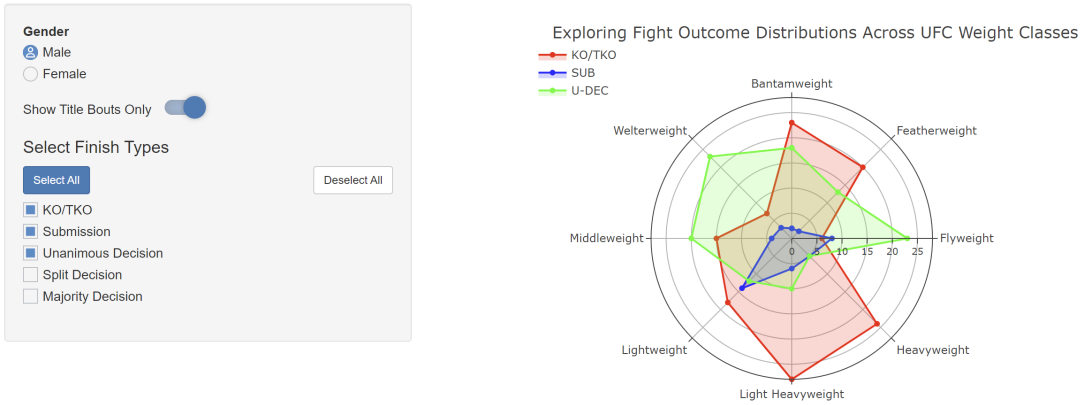

Figure 3.2: Idiom 2; some filters activated

The **second image** (Figure 3.2) illustrates the use of the **zoom feature**. By dynamically adjusting the radial axis range, the user focuses on a smaller subset of values, highlighting specific data points with lower fight counts. This interaction makes it easier to analyze patterns or trends in weight classes with fewer finishes, reducing visual clutter and improving interpretability.

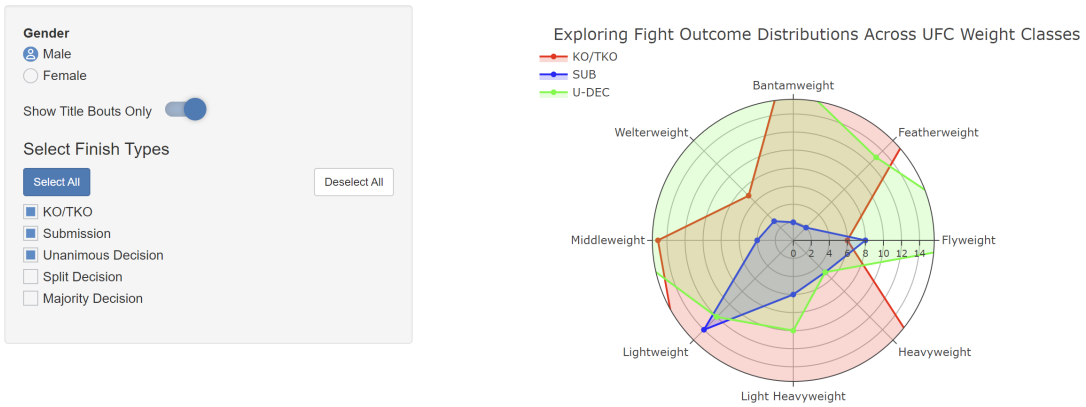

Figure 3.3: Idiom 2; some filters activated, zoomed-in with hover tooltip

4. Idiom 3 - UFC Fight Performance Heatmap

4.1 Task Definition

The primary goal of this idiom is to explore trends in UFC fight analytics by visualizing the relationship between the winner's age, nationality, and the average number of rounds fought. The heatmap provides an interactive representation where the intensity of shading indicates the average number of rounds across different countries and age groups. Users can filter the data by year range, gender, weight class, and fight location to gain deeper insights into performance trends. This visualization helps answer key analytical questions such as:

- How does the winner's age influence the average number of rounds in a fight?
- Which countries tend to have fighters engaging in longer or shorter bouts?
- Are there observable trends when filtering by gender or weight class?
- How have fight patterns evolved over time?

The interactive heatmap is designed for exploratory analysis, enabling analysts and fans to uncover patterns in fight performance and outcomes across multiple demographic dimensions. It supports comparative analysis across countries and age groups, offering insights into fight strategies and athlete longevity in the UFC.

4.2 Data Abstraction

The dataset used in this visualization includes attributes relevant to UFC fight performance and outcomes:

- *Winner's Age*: Numerical attribute representing the age of the winning fighter.
- *Country*: Categorical attribute representing the nationality of the winning fighter.
- *Number of Average Rounds*: Numerical attribute representing the average number of rounds completed in the fight.
- *Weight Class*: Categorical attribute indicating the weight division of the fight (e.g., Lightweight, Heavyweight).
- *Gender*: Categorical attribute indicating whether the fighters are male or female.
- *Year*: Numerical attribute representing the year in which the fight took place.

- *Fight Location*: Categorical attribute indicating the country where the fight occurred.

The data is aggregated based on the winner's age and country, with the intensity of the heatmap representing the average number of rounds completed for each combination. Filters for year range, gender, weight class, and fight location allow users to refine their exploration. This abstraction supports a comprehensive analysis of fight performance trends across multiple demographic and competitive dimensions.

4.3 Task Abstraction

This idiom is aligned with the actions of **discover** and **compare**.

Through **discovering**, users can identify patterns in fight performance by exploring how the winner's age, country, and number of average rounds correlate. It allows analysts to uncover which age groups and nationalities tend to participate in longer or shorter bouts and how performance varies across different demographic filters.

On the other hand, by **comparing**, users can examine differences between various countries, weight classes, and genders, identifying how fight patterns shift across these dimensions. For instance, analysts can compare whether older fighters engage in longer bouts or if certain countries tend to have more extended fight durations on average.

4.4 Visual Encoding

The visualization employs a **heatmap** to encode the data:

- **X-axis:** Represents the winner's age, allowing users to explore how fight outcomes vary across different age groups.
- **Y-axis:** Represents the country of the winning fighter, providing a comparative view of fight performance by nationality.
- **Color:** The color intensity represents the average number of rounds fought, with darker shades indicating a higher average number of rounds.
- **Interaction:** Filters for year range, weight class, gender, and fight location allow users to dynamically refine the visualization and focus on specific subsets of the data.

Figure 4.1: UFC Heatmap showcasing winner's age, country, and average rounds

This encoding allows users to effectively identify patterns and trends in fight performance, such as how the average number of rounds varies across age groups and countries, while supporting comparative analysis across different demographic filters.

4.5 Interactions

The heatmap visualization provides a range of interactive features implemented using Shiny and Plotly, enhancing user engagement and analytical depth. These interactions allow users to explore fight performance trends across age groups, countries, and other dimensions in a highly customizable and intuitive manner. Below are the key interactive elements:

4.5.1 Year Range Slider

Description: A slider allows users to filter the fights by the year in which they took place.

Purpose: This filter enables exploration of fight trends over time, allowing users to focus on specific periods of interest.

Implementation: The `sliderInput("timespan", ...)` element sets a range for filtering the dataset based on the fight year. The server-side filter applies this using the `filter` function within the reactive expression `filtered_data()`.

4.5.2 Country Grouping Selector

Description: A dropdown menu allows users to group the data by different categories, such as country or fight finish type.

Purpose: This feature supports comparative analysis across regions or finish types.

Implementation: The `selectInput("heatmap_var", ...)` dynamically adjusts the grouping variable using `!!sym(input$heatmap_var)` in the `grouped_data()` reactive expression.

4.5.3 Weight Class Filter

Description: A dropdown menu enables filtering the dataset by the selected weight class.

Purpose: This filter allows users to focus on specific weight divisions and compare their fight patterns.

Implementation: The `selectInput("weightclass", ...)` modifies the dataset by applying a conditional filter within the `filtered_data()` reactive expression.

4.5.4 Gender Filter

Description: A radio button lets users filter data by gender: All, Male, or Female.

Purpose: This interaction allows users to compare fight performance between male and female fighters.

Implementation: The `radioButtons("gender_filter", ...)` modifies the dataset by filtering the `Gender` column within the `filtered_data()` reactive expression.

4.5.5 Fight Location Filter

Description: A text input field allows users to filter fights by the country where the fight occurred.

Purpose: This feature provides insights into location-specific fight trends and patterns.

Implementation: The `textInput("fight_location", ...)` is processed in the server using a conditional filter with `grepl` to allow partial matches on the fight location.

4.5.6 Age Category Binning

Description: A checkbox allows users to toggle between continuous and binned age categories on the heatmap.

Purpose: Binning simplifies the visualization by grouping fighters into age ranges, improving readability.

Implementation: The `checkboxInput("show_binned", ...)` triggers the use of the `binned_data()` reactive expression, which applies the `cut` function to create age bins and groups data accordingly.

4.5.7 Heatmap and Color Encoding

Description: The heatmap uses color intensity to represent the average number of rounds fought.

Purpose: Darker colors indicate a higher number of average rounds, enabling users to identify areas with longer fights visually.

Implementation: The color gradient is controlled by the `scale_fill_gradient` function in `ggplot2`, using a continuous color scale from "lightskyblue" to "navy."

4.5.8 Hover Tooltips

Description: Tooltips appear when hovering over heatmap cells, providing additional details about the data points.

Purpose: Tooltips provide detailed information for more precise analysis without overwhelming the visualization with labels.

Implementation: The `ggplotly()` function from the `plotly` library automatically generates tooltips showing the winner's age, country, and the average number of rounds.

These interactive features provide a flexible and comprehensive way to explore UFC fight data, empowering users to discover patterns and compare performance across multiple dimensions dynamically.

The **first image** (Figure 4.2) represents the default setup of the heatmap visualization. The user has selected the full year range from 2010 to 2023, grouped the data by `Country`, and included all weight classes and genders without additional filters. This configuration provides a broad overview of UFC fight performance, visualizing how the average number of rounds varies based on the winner's age and nationality.

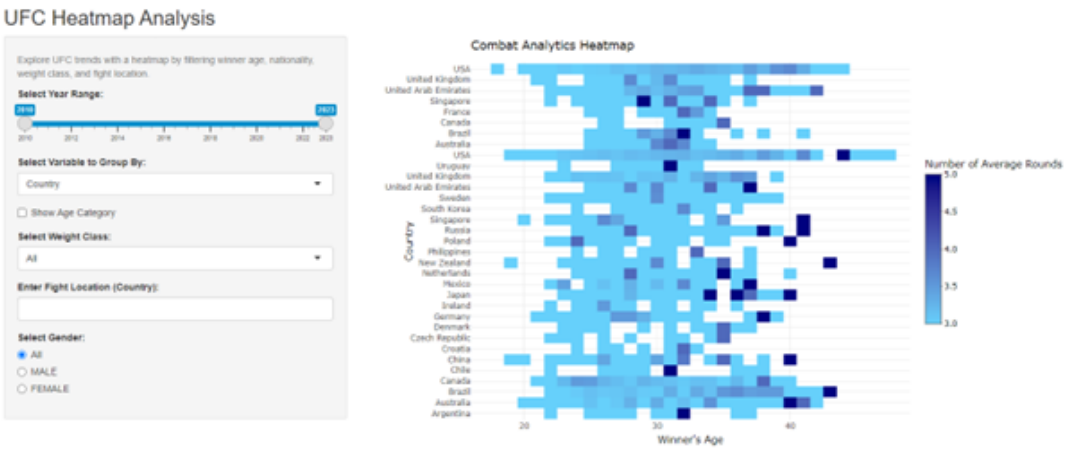

Figure 4.2: UFC Heatmap with default filter settings applied.

The **second image** (Figure 4.3) shows a filtered setup where the user has restricted the year range to 2013-2020, selected the `Lightweight` division, and filtered for male fighters only. The heatmap highlights fight patterns within the lightweight category, showing how the average number of rounds correlates with the winner's age and nationality. Darker shades on the heatmap indicate a higher average number of rounds fought.

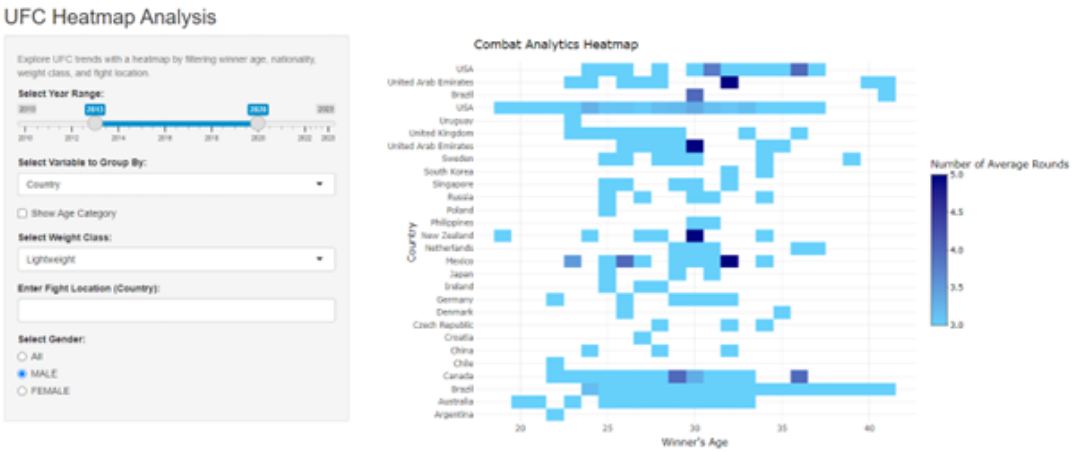

Figure 4.3: Filtered UFC Heatmap focusing on male fighters in the Lightweight division between 2013 and 2020.

These interactive filters allow users to explore detailed fight performance trends by focusing on specific dimensions such as weight class, gender, and time periods, providing a more granular analysis of fight outcomes.

5. Idiom 4 - Geographical Distribution of UFC Fights

5.1 Task Definition

The primary goal of this visualization is to explore how UFC events are distributed geographically across locations, focusing on the overall frequency of events as well as the prevalence of underdog victories. The analysts can examine the popularity of UFC events in different regions and the frequency with which unexpected outcomes occur. The interactive features enable users to refine their exploration, addressing the following questions:

- Which regions host the highest number of UFC events?
- How do fight frequencies compare between locations?
- Which regions have the highest percentages of underdog wins?
- Are there significant differences in the distribution of events or underdog performance based on the gender of the fighters?

This idiom empowers analysts to uncover patterns and trends in the geographic distribution of UFC events and results.

5.2 Data Abstraction

The dataset used for this visualization is the one explained in Section 1.1. The key columns used for this task include:

- *Country*: String representing the name of the country where the fight took place.
- *Location*: String representing the specific location (city/state) of the fight. This attribute is used to extract the state for fights held in the United States.
- *Gender*: Categorical attribute representing the gender of the fighters (either MALE or FEMALE).
- *Winner*: Categorical attribute representing the corner that won the fight (either Red or Blue).
- *BlueOdds*: Numerical attribute representing the odds for the blue fighter.
- *RedOdds*: Numerical attribute representing the odds for the red fighter.

In addition to the fight data, geographical data is used for visual mapping. This data is contained in two GeoJSON files:

- **countries.geojson**: Contains polygon data for each country, allowing us to map UFC fight statistics globally.
- **us_states.geojson**: Contains polygon data for each US state, enabling mapping of fight statistics at the state level within the United States. This data is particularly useful because most UFC fights take place in the United States.

After processing the fight data and the geographic data separately, they are merged to allow for spatial visualization. Before merging, the following preprocessing step was performed:

- Renaming the USA: The fight dataset originally used "USA" to refer to the United States of America. To ensure consistency with the geographic data in the GeoJSON file, "USA" is renamed to "United States of America" in the fight data before merging.

The resulting merged datasets associate fight information with the respective geographic regions (countries and US states) in the GeoJSON files, making it possible to calculate statistics that will be displayed on the interactive map.

Furthermore, measurements have been calculated to support the analysis based on the dataset. These measures include:

- Total Fights: The total number of fights in each country or US state, which is used for normalization when calculating the percentages.
- Fight Percentage: The ratio of the total number of fights that occurred in each country or US state to the total number of fights in the dataset. It is calculated by dividing the number of fights in a region by the total number of fights and multiplying by 100 to get the percentage.
- Total Underdog Fights: The total number of underdog fights in each country or US state. An underdog is determined based on the odds: if the red fighter's odds are greater than the blue fighter's odds and the blue fighter wins, or if the blue fighter's odds are greater and the red fighter wins, the bout is counted as an underdog victory. The total number of underdog fights is used for normalization when calculating the underdog fight percentages.
- Underdog Fight Percentage: The percentage of fights in which the underdog won. The underdog percentage is calculated by dividing the number of underdog victories by the total number of fights in that region and multiplying by 100.

5.3 Task Abstraction

This task focuses on analysing UFC fight events to uncover **trends** and **outliers** in their geographic distribution and the prevalence of underdog wins. Also, to compare two different regions by their **similarities**. By examining the proportion of events held in each region, it highlights areas that are dominant in the staging of events and identifies regions with higher loss rates, providing insight into the overall and regional dynamics of the fights.

These analyses help **discover** patterns such as the concentration of fights, variations in loser wins and changes in event locations. This approach provides a clearer understanding of the evolution of the UFC's geographic footprint and the relative importance of different regions in its global operations.

5.4 Visual Encoding

A **choropleth map** is used in the visualization to encode the data:

- **Regions**: Geographic areas, such as countries and U.S. states, are represented as polygons on the map.
- **Color**: The color intensity of each region encodes the percentage of UFC fights held there, with darker shades representing higher percentages.
- **Popups**: Hovering over a region displays specific statistics, such as the total number of fights, the percentage of events and, optionally, the number and percentage of underdogs in that area.
- **Legends**: Legends are included to indicate the color coding scale, providing a reference for interpreting the color gradients representing the percentage of fights.

This encoding lets the user easily identify the regions with the most UFC activity, compare the distribution of fights between areas, and assess patterns of underdog wins in different locations.

5.5 Interactions

The choropleth map visualization offers several interactive features that enhance user engagement and depth of analysis. These interactions allow users to explore statistics related to fights in countries and states in a highly customizable and intuitive way. The main interactive elements are presented below:

5.5.1 Gender Filter

Description: A radio button enables users to filter the data according to the gender of the fighters.

Purpose: This filter allows users to compare the distribution of fights and underdog statistics between genders, uncovering possible trends or patterns specific to male or female fighters.

Implementation: The 'input$gender' value dynamically filters the dataset, ensuring only fights matching the selected gender are included in the map.

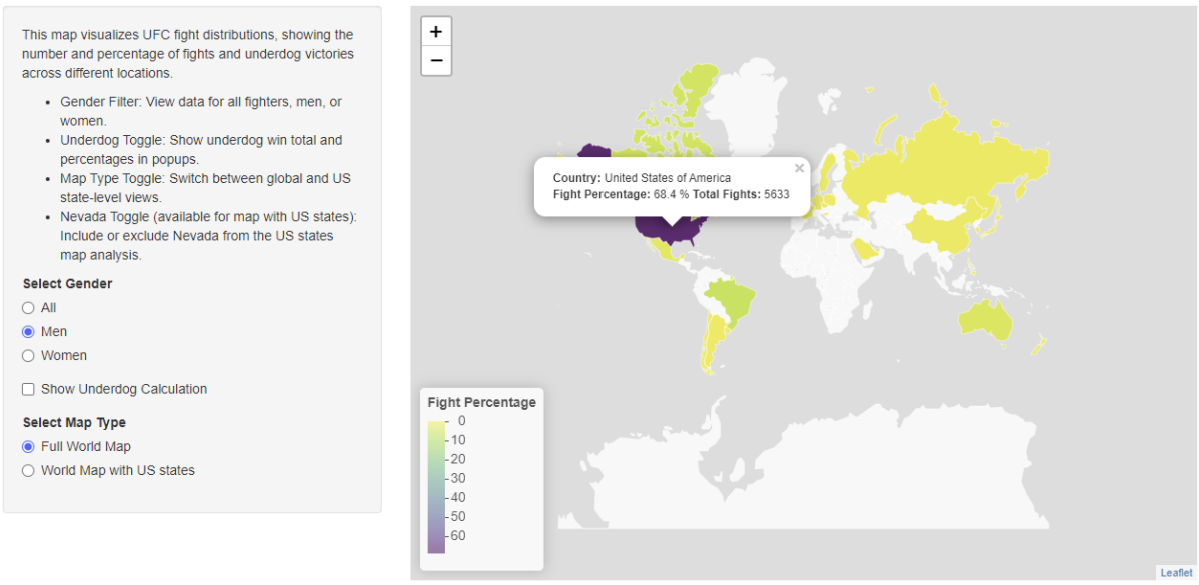

Figure 5.1: Idiom 4, filtering for *Gender* "Men".

5.5.2 Underdog Calculation Toggle

Description: A checkbox allows users to enable or disable the display of underdog fight statistics. When enabled, map popups display data specific to underdog fights, such as the percentage of underdogs.

Purpose: This feature focuses the analysis on underdog outcomes, helping users explore regions with high or low upset rates in UFC events.

Implementation: The value 'input$show_underdog' activates the conditional display of results obtained after calculating underdog statistics in popups.

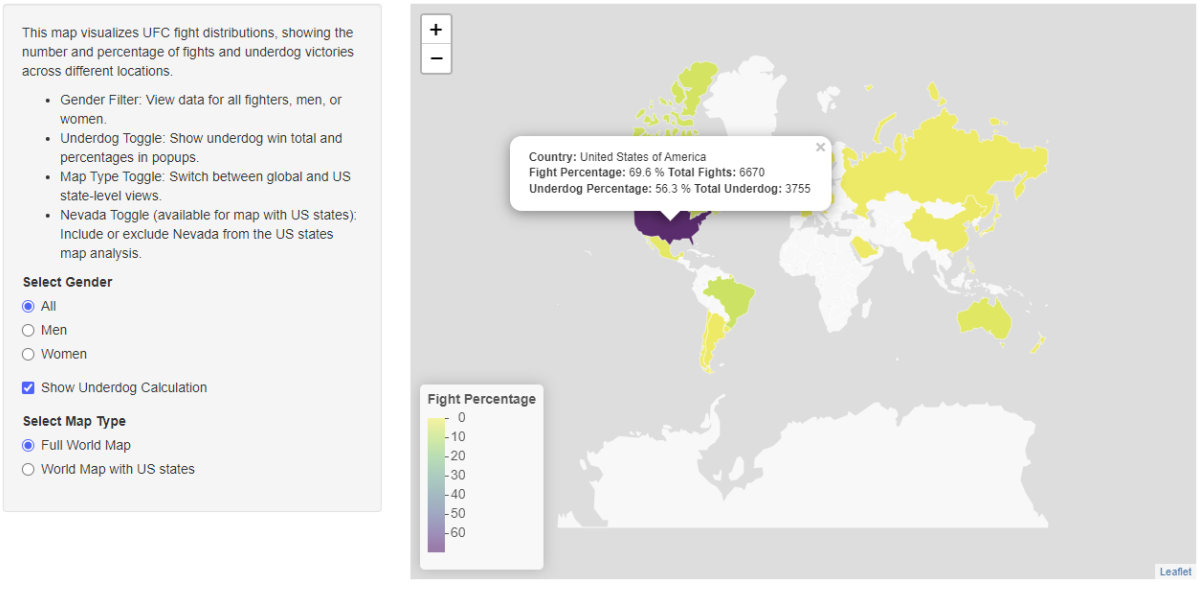

Figure 5.2: Idiom 4, enabling *Underdog Calculation*.

5.5.3 Map Type Selector

Description: A radio button selector allows users to switch between two map views: 'Full World Map' and 'World Map with US States'.

Purpose: This feature offers flexibility in geographic focus, allowing users to explore global trends or incorporate the states of the US in order to observe patterns better.

Implementation: The 'input$map_type' value determines the map representation logic. The map switches seamlessly between the two views, recalculating the metrics. When 'World map with US states' is selected, an additional checkbox ('Include Nevada in the Map') appears, providing more fine-grained control over the inclusion of state-level data.

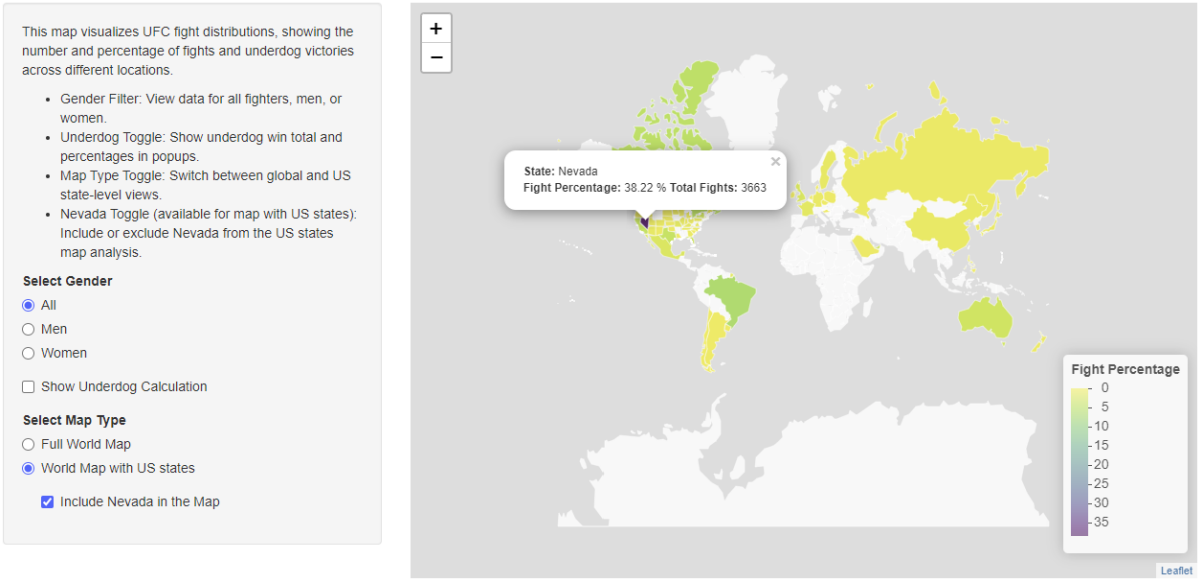

Figure 5.3: Idiom 4, selecting the *Map Type* "Word Map with US states".

5.5.3.1 Include Nevada Option

Description: A checkbox, conditional on the 'World Map with US States' view, allows users to include or exclude Nevada from the analysis.

Purpose: This option allows for more precise control of the analysis, given the high number of UFC fights located in Nevada.

Implementation: The 'input$include_nevada' value determines whether fights occurring in Nevada are included in the dataset. If unchecked, fights in Nevada are excluded from the visualization and related calculations.

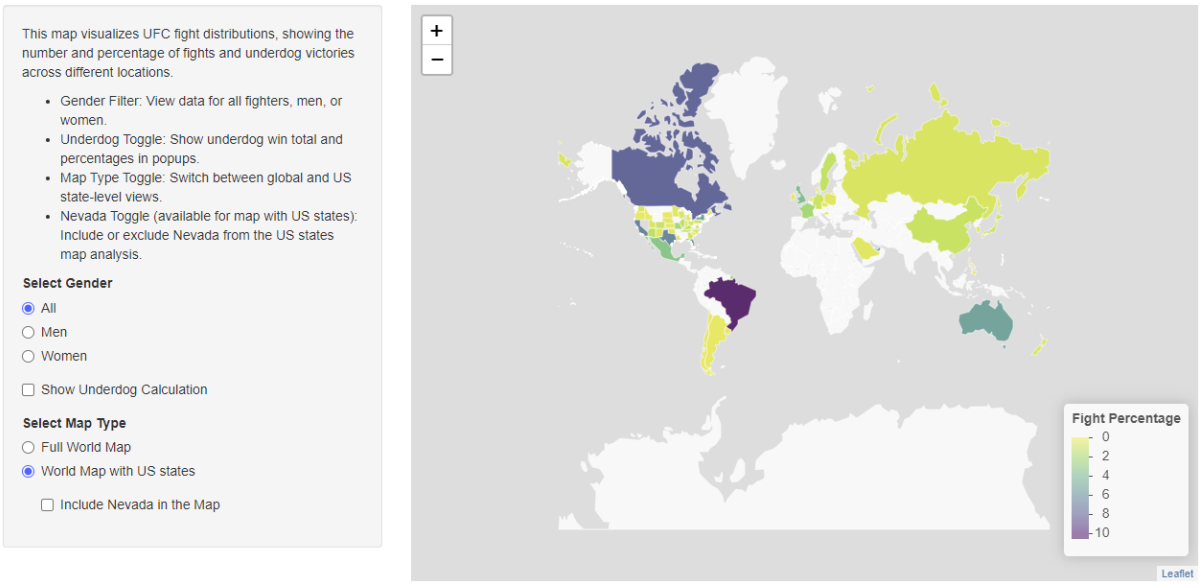

Figure 5.4: Idiom 4, selecting the *Map Type* "Word Map with US states" and disabling the inclusion of the state of Nevada.

5.5.4 Hover-Over Popups

Description: Hovering over a region on the map displays detailed information in a popup.

Purpose: This feature allows users to explore specific fight-related statistics for each country or state in an intuitive and immediate way.

Implementation: The popup displays dynamic content based on the selected filters and toggles. Its content includes:

- Region Name: The country or state name.
- Total Fights: The total number of fights held in the region.
- Fight Percentage: The percentage of total fights occurring in the region relative to the global total.
- Total Underdog Fights (if enabled): The total number of fights in which the underdog fighter won in the region.
- Underdog Fights Percentage (if enabled): The percentage of fights in the region that resulted in an underdog victory.

5.5.5 Dynamic Legend Updates

Description: The color legend dynamically updates to match the active map view and data range.

Purpose: This ensures that the visualization is visually consistent and accurately represents the displayed data.

Implementation: The legend scale adjusts based on the maximum value of the fight percentage in the dataset.

5.5.6 Data-Driven Map Rendering

Description: The map dynamically renders to reflect all active filters, toggles, and user inputs.

Purpose: Provides a seamless and responsive user experience that adapts to changing analytical needs.

Implementation: Filters and toggles dynamically trigger reactive calculations and updates to the map layers, ensuring that all displayed maps include only the data relevant to the user inputs. Smooth transitions between views allow users to explore different analyses seamlessly, maintaining an uninterrupted and intuitive experience.

6. Instructions

The tool is published under https://davissiemens.shinyapps.io/final/.

To run the app locally, please run the *app.py* file, locally in *RStudio* as a Shiny App. Before running the app, please apply the below install commands.

```
install.packages("leaflet")
install.packages("sf")
install.packages("dplyr")
install.packages("readr")
install.packages("tidyr")
install.packages("stringr")
install.packages("ggplot2")
install.packages("plotly")
install.packages("fmsb")
install.packages("shinyWidgets")
```

To navigate to the different views, please use the corresponding tabs shown in the Figure 6.1. The instruction on how to use each interaction, and possible use cases can be found in the Chapters of the Views.

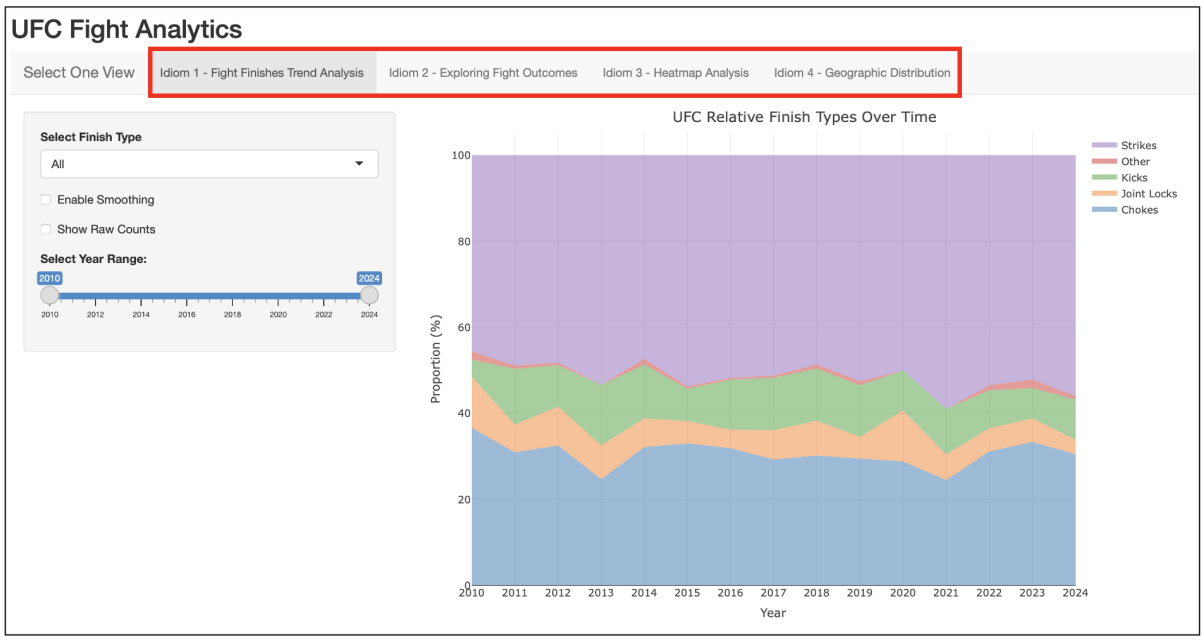

Figure 6.1: Here you can see the UFC Analytics App in full view. To navigate to the views 1-4, please use the corresponding tabs.

Bibliography

[1] Md Abbert. *Ultimate UFC Dataset (2010-2024)*. Accessed: December 28, 2024. 2024. URL: https://www.kaggle.com/datasets/mdabbert/ultimate-ufc-dataset.

[2] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015. ISBN: 9781498759717. URL: https://books.google.de/books?id=NfkYCwAAQBAJ.

[3] Ben Shneiderman et al. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 6th. Pearson, 2016. ISBN: 013438038X.