**Coding Data and Literacy Report:**

**A very brief explanation of your project ("I designed an app to that matches cheese types and wine types..."),**

**its current state (complete/incomplete, working/broken, buggy/unstable, etc.),**

**information about *how you developed your project*, including key *concepts* that you engage with during this semester.**

**and overall, how you strengthened your code literacy during the semester as you developed your project**

**Report:**

Project explanation:

I have placed together a range of data visualisations that communicate various cricketing data points. These data points have been gathered from a data set on Kaggle by user Dheeraj Mukati and provide a large range of insights regarding the top 86 batsman from around the world. This data set combines runs from all 3 formats of cricket including T20, ODI and Test matches as I am concerned with the cumulative runs scored for these players and how they compare country by country. Some of the metrics recorded within the data set are a players strike rate, high score, balls faced, the amount of 50s and 100s they have scored, cumulative runs and most importantly their average.

Aim:

The 5 visualisations depict the difference between overall runs and batting averages to determine which country has been the most dominant within these top 86 batsman and which batsman have performed the best. There isn’t a range of databases out there that compares cumulative runs vs averages, so I wanted to create a user-friendly visualisation that is easy to interpret, displaying this data in a map (figure 2). Furthermore this data was easily able to be quantified with a columns table in which countries were placed from highest to lowest (in regards to average of the countries average and overall runs). This table has been placed below and colour coded in a dashboard so that all the data is easily recognisable(figure 2). My other visualisations seek to communicate which individual player has been most dominate over all 3 formats and the difference between a players average and their high score (figure 3). Comparing this has helped to conclude if scoring a large total correlate to overall long-term consistency and what situations has this happened in. This has also helped to envision the difference in playing style as different formats describe different narratives (i.e. shorter formats leading to lower high score totals).

Overarching all of this research is the functional art flow chart depicted by Cairo as his key questions have helped to refine my work. Circumnavigating throughout this chart i have built up a concise understanding of my dataset and the specific question I want answered from this research:

*What data do you have?*

* Runs regarding the top 86 modern batsman, strike rates,averages high scores, balls faced, 50s, 100s, locations, etc.

*What do I want to know about my data?*

* What countries have best performed and how the switch between displaying and average and cumulative runs flip the narrative

*What do you see and does it make sense?*

Furthermore, some of the choices of visualisations have stemmed from articles regarding effective data reporting. This refers to the addition of vertical bar charts underneath the map visualisations as it is highly recommend quantitative data is presented incrementally and linearly (see figure 2). This was done by using the ‘sort by highest to lowest’ tool

Method:

Firstly I translated the data from the excel file and placed it into a Tableau folder quickly realising that the location given was not integrable by Tableau (i.e Tableau didn’t recognise terms such as IND and AUS as real locations). Therefore, I filtered out these keywords to organise the players by their countries. I then used the function ‘=(IF(ISNUMBER(SEARCH("IND",B2)),"India",""))’ which searched for the keyword of ‘IND’ in the ‘B’ column and translated into a more recognisable value. This function was then used for each of the 9 countries and Tableau would be able to place a geographical location on it. This then allowed me to map my data onto a geographically accurate map with the participating countries highlighted. I then manipulated various other parts such as a Figure 1.01
progressive colour tool to show the variation in the number of runs/averages. I then used the ‘tool tip’ function bar to create a more approachable hover state highlighting the 2 important points of the country and the data point. Originally the hover state involved 4 decimal places describing the average and took up too much of the screen, so it was important that a more digestible interface was Chart

Description automatically generated with medium confidenceused.

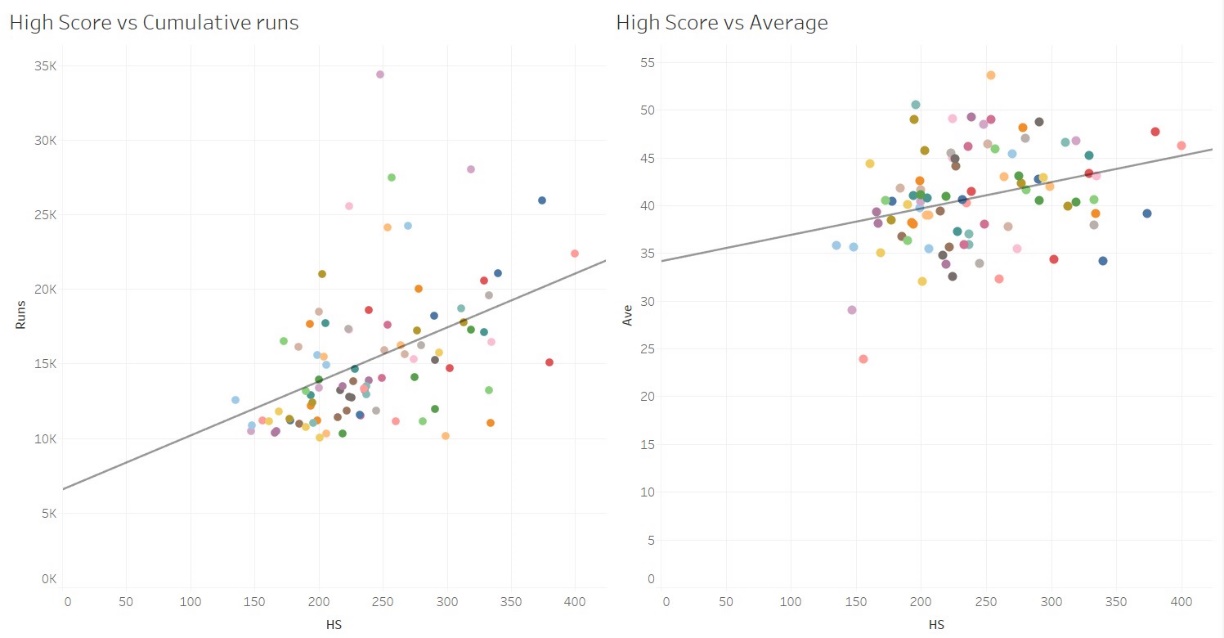
Figure 1

Figure 2

For the insights that reflect the playing abilities of the top 86 batsman I chose to use a scatter plot involving the 3 metrics the the player, (average runs) and

Conclusions:

Overall, these insights have confirmed



Chart, scatter chart

Description automatically generated

How it was developed and concepts coding concepts:

**Video:**

**Coding Literacy examples:**

* Excel spreadsheet using the *if* function to convert abbreviated location into a location Tableau could recognise. =(IF(ISNUMBER(SEARCH("SL",B2)),"Sri Lanka",""))
* =(IF(ISNUMBER(SEARCH("IND",B2)),"India",""))
* Tooltip changes
* Changes for geographical roles

**References:**

*Most Runs in International cricket*. (n.d.). Www.kaggle.com. Retrieved November 7, 2022, from https://www.kaggle.com/datasets/dheerajmukati/most-runs-in-cricket

**‌**