Infovis Project Final Report

Shiyi Huo, Rui Ma, and Mohamed Ghannam

Abstract— This is a project final report of group Alphago for UVic CSC 485D/SENG480D course in Spring 2018. It proposes a dashboard visualziation for exploring relationship between i) money received and team performance and ii) transfer fee and team performance for football. Idioms used are validated, scenarios of use are given, and pros and cons of design and implementation are discussed. Usability test is conducted and the feedback is gathered. The video link and evaluation document is given in appendix. You can also find them in GitHub repository.

Index Terms— football, visualization, money received, transfer fee, team performance, interactive, dashboard, bar chart, stacked bar chart, scatter plot, animation

INTRODUCTION

This report contains project introduction, approach (user, data and task), related work, design, implementation, evaluation, findings, discussion and future work.

1 PROJECT INTRODUCTION AND MOTIVATION

1.1 Main Idea

Football is one of the most popular sports in the world, with millions of fans following and supporting their teams. Every country has its own respective league, and noticeably, there are four-major leagues that draw the most attention [1]. And every league consists of 18 to 20 teams.

Over the last couple of years, there has been a huge increase in the amount of money in the game from TV deals. Television companies prepare bids for each league association to buy the rights of broadcasting the games. As football gains even more popularity, these TV deals can reach billions today. This influx of money has caused football fans to question the effects it has on the game. The reason why money is an important part of the game, is because teams have to buy new players, renew contracts, and pay wages. One effect can be positive, where with this increase in money, more teams will be able to build and improve. Thus improving competitiveness between clubs and giving fans more excitement. Others believe it can have a negative effect, where only the rich will get richer and there will be less competitiveness. Our visualization is to examine whether football has simply become a game dominated by money, where teams that make the most money and spend the most are guaranteed to improve? This project will make it easier to explore what effects money has had on football.

1.2 Goal

With the tagline of "pay, players, and performance all in one dashboard", we want to create a football visualization that is easy to understand and use. Our goal is to help user find the correlation between money received and performance, and transfer fee and performance to see the effects of money on football. We plan to visualize the breakdown of the money that teams receive from the TV deals for each league, along with the amount of money that teams have spent on transfer fees. A transfer fee is the amount that a team pays to buy a new player. For the performance metric, we will visualize the amount of goals that teams have scored and conceded. The goal can be more clearly described using user tasks (see 2.2).

1.3 Potential Impact

Our visualization has the potential impact of verifying conjectures that football fans may have before using this visualization. There are users who will believe that the increase of money has affected football negatively as only a small number of teams benefit and improve from that increase, others may think it has had a positive impact. With our visualization it will become easier to see if these initial beliefs are true or false, since they do not need to go through dozens of tables to come up with a conclusion. The visualization can also offer tactical insight to users. For example, by looking at a team's performance, users can identify what positions the team needs to reinforce. They can then see if the team buy players in these positions, and if there are improvements in performance. In summary, our visualization can answer many questions about the impact of which money has had on the game; and allow users to explore initial hypothesis they may have about the topic.

2 APPROACH

2.1 User

The target audience of our visualization is football enthusiasts. For example, football fans who are interested in examining the effects of expenditure and player transfer on competitiveness of teams (within each league). Since users may not have sufficient visualization knowledge, we tend to avoid using complex idioms.

2.2 Task

Tasks are refined for easier usability test evaluation after implementation is done. There are 2 main tasks (task 1 and 2) and each task has its own subtasks. Task 1.1, 1.2, 1.3 are subtasks of task 1 while task 2.1, 2.2 are subtasks of task 2. Note that task 1.3 and 2.2 are the same, but since they are in different context, different interaction idioms are used (see 4.2.6). Task hierarchy is listed below.

For each league:

1. Examine relationship between money received and team performance

    1.1 Find teams who received the most and the least amount of money according to breakdown.

    1.2 Examine relationship between money received and the position where a team finishes in the standings.

    1.3 Find the correlation between number of goals scored vs number of goals conceded to evaluate a team's performance

2. Find the correlation between player transfer cost vs team performance

    2.1 Find the transfer fee breakdown among player roles for each team

    2.2 Find the correlation between number of goals scored vs number of goals conceded to evaluate a team's performance

Vis design is closely focused on tasks presented above, so users will be able to perform tasks more easily with our vis.

Also note that there are two metrics for evaluating individual team performance: team ranking for one season and number of goals scored vs conceded. The former is a general metrics and it considers accidental factors like injury, while the latter is a detailed metrics and it is more tactical. But both of them are important.

2.3 Data

For each league, there are basically three tables (datasets) [12][13][14].

Table 1 is called money received table. It contains amount of money received and team performance data, and it is used for task 1 (see 2.2). There are 6 attributes in this table: year/season, position/ranking of a team, team name, total money received for one team one season (in million £), number of goals scored and number of goals conceded. Within the dataset, year is a temporal attribute, position/ranking of a team is a sequential (ordinal) attribute, team name is a categorical attribute, total money received, number of goals scored and conceded are all quantitative attributes.

Table 2 and 3 are both of player transfer data, and they are used for task 2. Table 2 is called total transfer fee table. There are 3 attributes in this table: year/season, team name, and total transfer fee for one team one season (in million £). Within the dataset, year is temporal, team name is categorical, and total transfer fee is quantitative.

Table 3 is called detailed transfer fee table. There are 6 attributes in this table: year/season, team name, player arrival, player role, moving from, individual player transfer fee. Within the dataset, year is temporal, individual transfer fee is quantitative, and all other attributes are categorical.

3 Related Work

There is one paper [2] found for visualizing player transfer data, and there are some related vis found. The paper talks about two idioms for visualizing player transfer data: Chord diagram and force-directed placement of node-link diagram. Most football vis only focus on one task while our project focus on two main tasks. So the dashboard layout we proposed is new for football vis.

3.1 Existing Solutions

Existing solutions are presented for different subcomponents of our vis.

3.1.1 Comparison Between Teams

One visualization (Fig. 1) focuses on all teams within English league over a number of seasons, and its goal is to compare team related statistics [3]. It is a dashboard which consists of a scatter plot, a bar chart and a map. A drop-down menu is used for users to select the statistic that they want to look at, and a slider is used to change seasons. Issues with this vis is 1) lack of legend for all idioms, users do not know what each color means; 2) discriminability issue within all idioms; 3) scalability of idiom is not carefully considered during design, so clustering happens within almost every idiom.

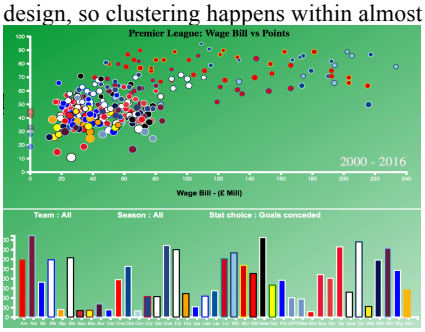

Fig. 1. Dashboard of Football Statistics

3.1.2 Team Performance Evaluation

One vis related to team performance is a box plot (Fig. 2) showing number of goals scored and conceded for two teams [4]. One pro of this idiom is that it includes the average line of number of goals scored. Cons of this idiom is that 1) we need to include 18~20 box plots which may cause scalability issue; 2) unaligned boxplot is not intuitive for comparison tasks.

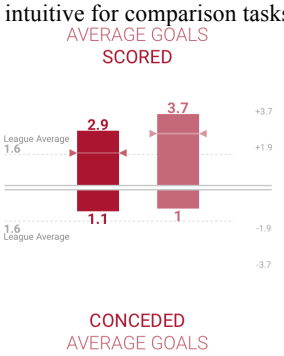

Fig. 2. Box Plot for Number of Goals Scored and Conceded

Another vis includes various idioms to show different aspects of goals scored and conceded for Liverpool in season 16/17 [5]. It shows number of goals scored distributed among players, types of goals conceded and when goals scored/conceded. It also uses line chart to show historical data of number of goals scored (Fig. 3). Using line chart for showing number of goals scored and conceded was discussed for this design and we conclude that since there are only four seasons of data, it is not very valuable to show a line chart.

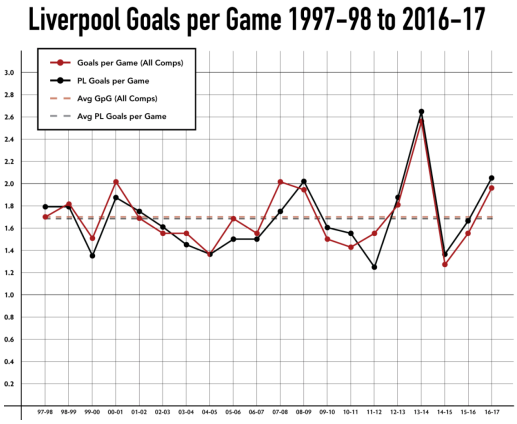

Fig. 3. Line Chart for Number of Goals Scored and other Statistics

3.1.3 Player Transfer Data

Visualizing transfer data is a major challenge of this project since transfer data is usually of high dimension. Different visualizations about transfer data will be discussed below and each vis has its own focus.

The most traditional vis for transfer data is to use a list/table (Fig. 4) to show player name, team from, team to and transfer fee [6]. Team badges are added beside team names to make vis more vivid and teams more distinguishable. Since transfer data is of high dimension, original design of idiom is also a table (see 4.2.5). But using list/table simply presents data and does not provide more insight.

Another vis focuses on showing large transfer fees for players. It uses a bar chart (Fig. 5) to show player name, season and corresponding amount of transfer fee [7]. Since player name could be long, it uses horizontal bar chart rather than usual vertical bar chart for better legibility. Our design adopts the use of horizontal bar chart as well (see 4.2.3).

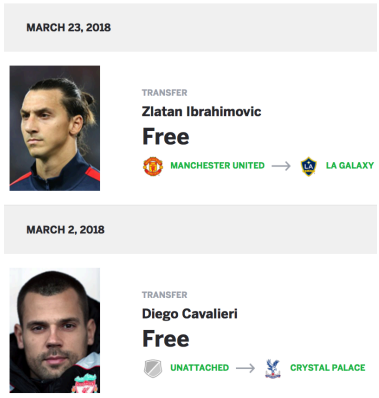

Fig. 4. Player Transfer Vis – Player List

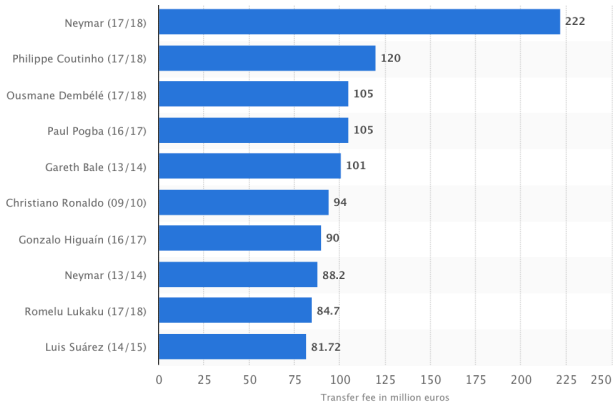

Fig. 5. Player Transfer Vis - Bar Chart for Top Transfer Fees

Geometry map (Fig. 6) uses point marks to denote team locations and line marks to denote paths of player transfer [8]. This particular vis does not show player name and individual player transfer fee. But this information could be integrated into map and be shown when a line is hovered. This approach will not be adopted for this project since geographic data is not a concern and this idiom requires a lot of screen space.

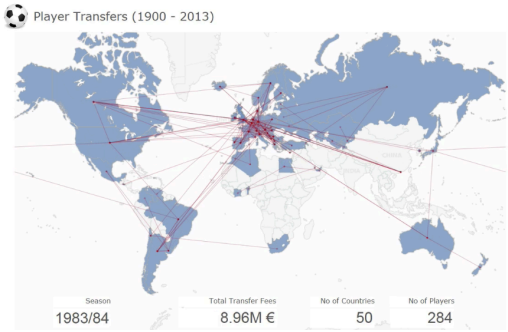

Fig. 6. Player Transfer Vis – Map

Node-link diagram (Fig. 7) uses team badges as point marks and directed edges as line marks [9]. But it is hard to avoid overlap between edges and it is hard to put much annotation on one edge. Force-directed placement may help a bit with the overlapping issue.

Chord diagram (Fig. 8) is more like a node-link diagram in radial layout but how to arrange data should be carefully designed [2]. It suffers from the same issue as node-link diagram but it is indeed a more creative idiom. Using Chord diagram requires data transformation and it will be discussed in detail in 4.2.5.

Radar chart (Fig. 9) is usually used when comparing competencies of individual player and average player [10]. It can be used together with individual player transfer fee as a detailed view for transfer data. However, since player competency evaluation is subjective and the data is hard to get, radar chart will not be used for this design.

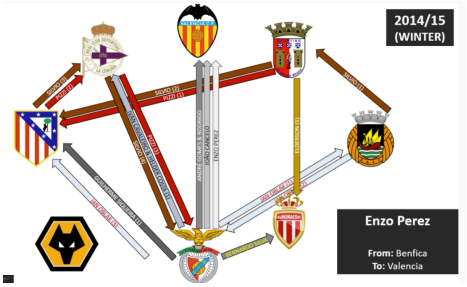

Fig. 7. Player Transfer Vis – Node-Link Diagram

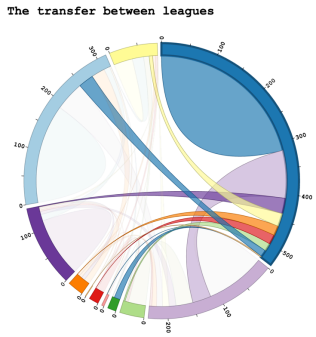

Fig. 8. Player Transfer Vis – Chord Diagram

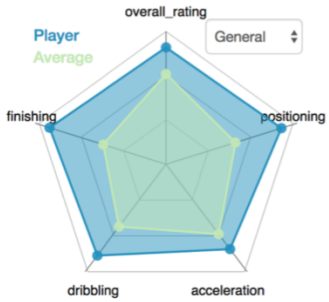

Fig. 9. Player Transfer Vis – Radar Chart

There is also a storytelling visualization (Fig. 10) for transfer data [11]. During storytelling, different tasks are targeted and various idioms are used. Individual idioms are interactive and user can change year range or team selection for specific idioms. It is a really interesting way to visualize transfer data.

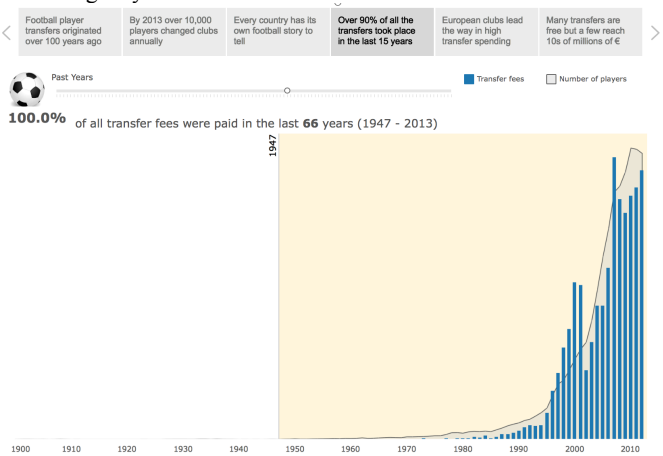

Fig. 10. Player Transfer Vis – Storytelling

3.2 Closest Design

The dashboard vis mentioned in Fig .1 is the closest thing to our design, but this vis has many drawbacks. We will also use dashboard as the overall approach and scatter plot for one idiom. We also look at some statistics over a number of years, however, we are looking at a smaller period of time; from the 2013-2014 season to the 2016-2017 season. The reason being, the increases of money from TV deals and sponsorships started in that year, and we run up to 2016-2017, because the current season is not over yet.

3.3 Contribution of Current Project

Since this project explores if money has had an impact on football, we need to look at more than one league to be able to come up with a sensible conclusion. Compared to most vis discussed above, our design uses tabbed dashboard to help users switch between the different leagues and tasks to see the appropriate visualizations. Interactions are added as users hover over elements in idioms to help them quickly identify related element between idioms. Besides, we explore the use of stacked bar chart for player transfer data, and reason that chord diagram is not the best vis of transfer data for the given dataset.

4 Design

4.1 Design Description

The design is a tabbed dashboard with 4 high-level tabs for leagues and 2 low-level tabs for tasks. Idioms are grouped by tasks and the user can select a league and a task to reveal related idioms. Recall that two main tasks proposed are 1) examine relationship between money received and team performance, and 2) examine relationship between transfer fee and team performance. See section 2.2 for more detail of tasks.

There are two idioms for task 1 (Fig. 11): a bar chart for money received vs team ranking and team name is on the left, and a scatter plot for number of goals scored vs conceded on the right. There are two idioms for task 2 (Fig. 12): a stacked bar chart for transfer fee vs team name on the left, and a scatter plot for number of goals scored vs conceded on the bottom. Within each bar of stacked bar chart, transfer fee is broken down by player roles. Also note that the scatter plot is shared by both tasks since both tasks are about team performance where the scatter plot visualizes detailed team performance.

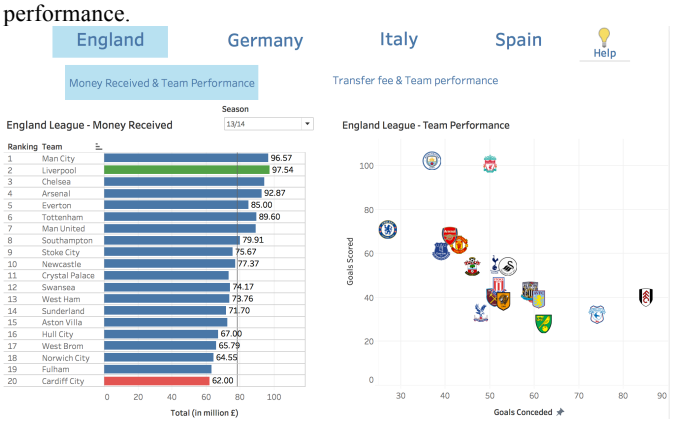

Fig. 11. Task 1 (Horizontal Bar Chart + Scatter Plot)

Within one task, interactions are added to help user relate idioms together. For example, for task 1, when user clicks a bar/team in bar chart, the corresponding team in scatter plot will be highlighted.

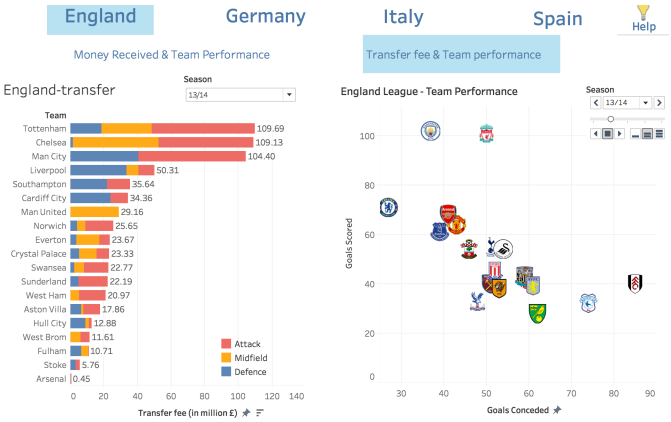

Fig. 12. Task 2 (Horizontal Stacked Bar Chart + Scatter Plot)

4.2 Design Options

4.2.1 Dashboard Hierarchy

At the very beginning, we proposed to include all four idioms within the same page of dashboard: two on the top and two on the bottom (Fig. 13). And we have a global selector to select league. The intention of this layout is that users can select the information as needed

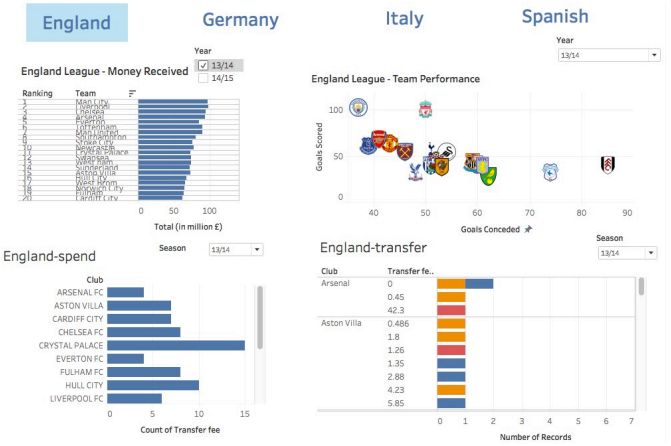

Fig. 13. Original Design: All Idioms in One Page

But there are several downsides for this layout.

1) Since all idioms are presented at once, there is no focus for the information presented. Users could get confused which idiom to look at, and they may end up doing something unexpected.
2) Both bar charts are used for all teams, so they take a lot of space, either horizontally or vertically. Besides, badges are used as scatters in scatter plot, so the scatter plot cannot be too small or the badges will not be distinguishable. So putting all idioms within the page is really clustered.
3) Since there are several idioms and some idiom comes with a filter to select year, there end up being several filters on dashboard. And it will also confuse user.

To get rid of the shortages of one-page dashboard layout, we decide to use tabbed dashboard [15]. Tabs in a dashboard is like a navigation bar for a website, and it navigates users between views. Besides, we choose every league as a high-level category and each task type as a low-level category. Note that tasks are the same for each league, so the user can select one league and then select a task to reveal related idioms.

4.2.2 Visual Encoding Idioms

There are 3 visual encoding idioms in total for the whole dashboard:

1) a horizontal bar chart for money received vs teams

2) a scatter plot for number of goals scored vs number of goals conceded
3) a horizontal stacked bar chart for total player transfer fee vs teams

Task 1 includes idiom 1 and 2, while task 2 includes idiom 2 and 3. Idiom 1 and 2 share the same filter for year, because prize money received for one season is related to the performance for that season. Idiom 2 and 3 have their own filters, because transfer fee for the current season will affect team performance in the future. Besides, both bar chart and stacked bar chart function more like overview, so they are positioned on the left since most users read from left to right. Details of each visual encoding idioms will be discussed in following sections.

4.2.3 Idiom 1: Horizontal Bar Chart

Idiom 1 (Fig. 14) is designed for task 1.1 (find teams who received the most and the least amount of money according to breakdown) and 1.2 (examine relationship between money received and the position where a team finishes in the standings).

Since final team ranking is a categorical key, team name is a categorical attribute associated with team ranking and money received is a quantitative attribute, we have several options: bar chart, line chart, pie chart and polar area chart. Besides, since the task is to look up and compare values between teams, bar chart is more appropriate. To better illustrate relationship between money received and ranking, we ordered ranking ascendingly. Bars with maximum and minimum money received are highlighted with green and red. The average line (vertical) is added so that user can hover on that line to see average money received for English league during 13/14 season, and he can easily see which team receives more than average.

Some team names are long (e.g. Atletico Madrid), so to show team names properly, we may need to use tilted text. Another problem with vertical bar chart is that since there are 18 to 20 teams, the chart will take much space horizontally. So to solve these two problems, we propose to use horizontal bar chart [17].

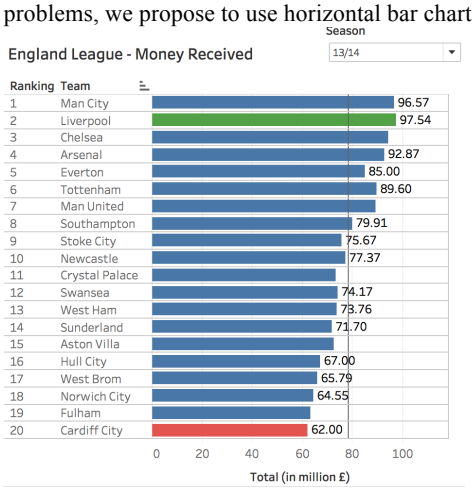

Fig. 14. Horizontal Bar Chart

4.2.4 Idiom 2: Scatter Plot

Idiom 2 (Fig. 15) is designed for task 1.3 and 2.2 (find the correlation between number of goals scored vs number of goals conceded to evaluate a team's performance). Since both number of goals scored and conceded are quantitative attributes, and the task is to find correlation between these two quantitative attributes, scatter plot is chosen for this task.

Team badges are used as icons for quickly identifying teams, and number of goals scored and conceded for each team will be shown when user hovers over one icon. It is hard to recognize a team when the icon is really small, so icon size is carefully examined. Because scatter size is relatively large, user cannot get accurate values by just looking at scatter plot. So hovering function is added to retrieve accurate statistics (Fig. 16). Besides, clustering or occlusion is likely to happen if there are several teams with similar performance, so axes are adjusted (they may not start from 0) to minimize occlusion.

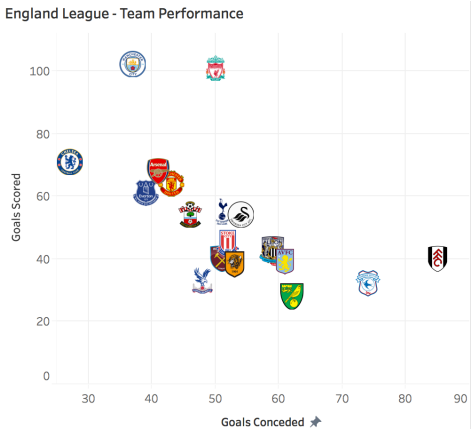

Fig. 15. Scatter Plot

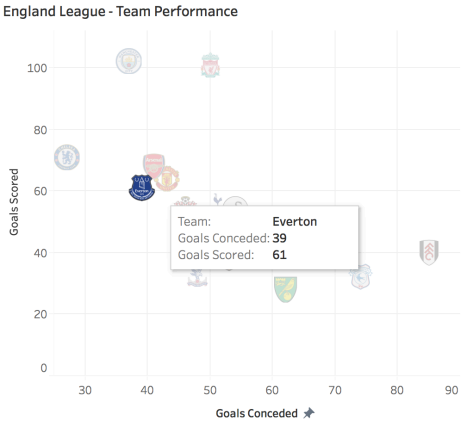

Fig. 16. Scatter Highlighted in Scatter Plot

4.2.5 Idiom 3: Horizontal Stacked Bar Chart

Idiom 3 (Fig. 17) is designed for task 2.1 (find the transfer fee breakdown among player roles for each team). Dataset involved for this task is table 2 and table 3. Since table 3 is of high dimension (see section 2.3), we had a hard time determining which idiom to use.

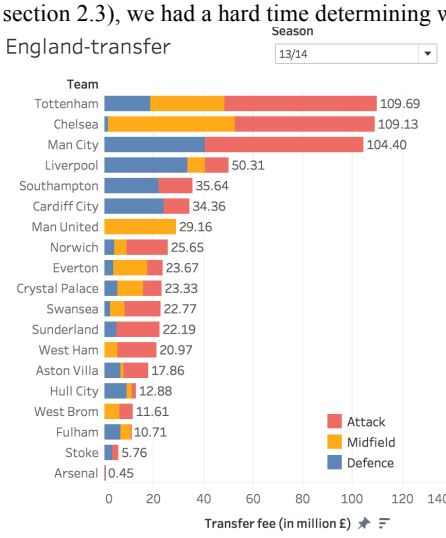

Fig. 17. Horizontal Stacked Bar Chart

Initial design is to use a table with 4 attributes: player name, player role, player from, and player transfer fee. The advantages of using table is that table is good for lookup task, and users are familiar with it and they can easily retrieve information they need. The disadvantage is that from the visualization design point of view, table is not creative and it cannot show much insight from data.

Current design is to use a horizontal stacked bar chart for transfer fee vs team name. Length of each bar encodes total money spent on player transfer for one team one season, and it is composed of money spent on different player roles. This task suffers the same problem of too many teams as task 1, so horizontal rather than vertical stacked bar chart is used. Within original data, player role attribute has more than 10 values. To make it fit into a stacked bar chart, player role is aggregated into 3 values as attack, midfield and defense, and now player role is a diverging attribute with midfield as a central point. These three values are carefully selected to match quantitative attributes in scatter plot. When user hovers on a portion (ie. attack) within one bar, related value will be highlighted (Fig. 18) [19][20]. Different from the horizontal bar chart (idiom 1), team names within this idiom is ordered by transfer fee for each team, and there is no ranking associated with it. Ranking for current year is not valuable anymore because users are now interested in player transfer fee for current year and team performance for the future.

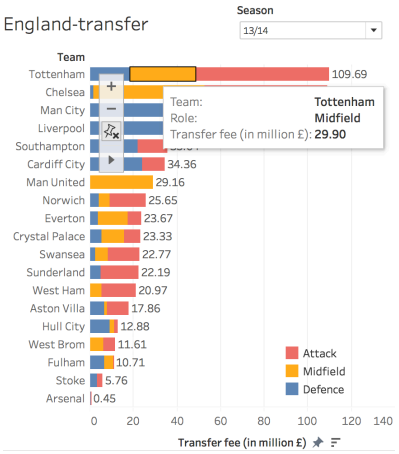

Fig. 18. Horizontal Stacked Bar Chart – Portion Highlighted

Chord diagram [2] was considered for this task because transfer data (table 2 and 3) is naturally network data. However, it is not adopted after careful validation. By using chord diagram, individual player transfer detail (player name, player role, player from, and player transfer fee) can be shown. Players could be transferred from a team in one league to a team in another league, so there will be about 80 data points around the circle, and each player transfer can be represented as an arc connecting two data points. But the problem with chord diagram is that many players does not come from teams within 4 leagues. One possible solution is adding one data point for "other" indicating the player is coming from other teams outside of four leagues. But since many players come from teams outside four leagues, it may end up with a chord diagram with many arcs starting from "other", while we want arcs to be approximately randomly distributed. Another challenge is that there may not be enough room for annotating arcs with player, position and price for the same receiving team, occlusion can occur, or annotation could be really small and so not legible.

4.2.6 Interaction Idioms

As for interaction idioms, there are

1) bidirectional highlighting between two idioms for task 1 through hovering (Fig. 19)

When user hovers on one bar in bar chart, the corresponding scatter in scatter plot will be highlighted; when user hovers on a scatter in scatter plot, the corresponding bar in bar chart will be highlighted. This interaction is added so that user can easily relate two idioms together and also see corresponding badge for a team name. Thus, no additional legend for scatter plot is needed. To find a badge for a team name using legend requires serial search, which is time-consuming. But in current design, corresponding badge is highlighted, so it pops up and finding it becomes preattentive [21].

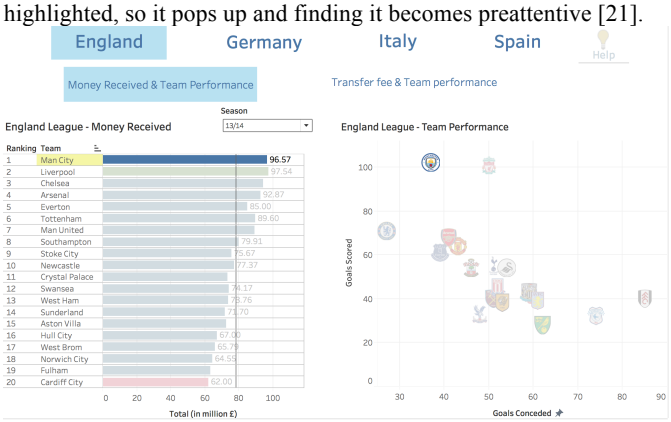

Fig. 19. Bidirectional Highlighting – Task 1

2) bidirectional highlighting between two idioms for task 2 through clicking (Fig. 20)

Bidirectional highlighting is also available for task 2. When user clicks on a team name in stacked bar chart, the corresponding scatter in scatter plot will be highlighted; when user clicks on a scatter in scatter plot, the corresponding bar in stacked bar chart will be highlighted. Note that clicking rather than hovering is used here since hovering is already used for highlighting each portion within one bar and two hovering actions may cause confusion. Other than that, we want user to use click to confirm their selection.

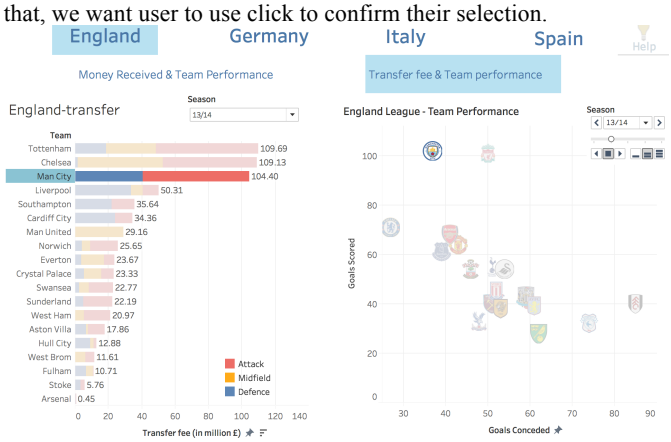

Fig. 20. Bidirectional Highlighting – Task 2

3) animation playing after play button is clicked for scatter plot in task 2

There is an animation menu on top right corner of scatter plot in task 2 and user can click play button to show animation of performance change across seasons for one team. After setting seasons for two filters, user can select a team in stacked bar chart and click play button (Fig. 21) to play animation (see animation in demo video). Animation is simulated by showing highlighted badge moving in scatter plot.

User is supposed to select a team and then play animation, so that the team selected pops up and its movement can be easily tracked. If user plays animation without selecting a team, he will see all badges moving at the same time. If he wants to follow a badge, he needs to perform serial search [21] every time the season changes. It

is not desired but note that the constraint, which is forcing user to select a team and then playing animation, is not implemented due to lack of time. The constraint is added into future work (section 9.2).

User can also set speed for animation. Since user can select season from dropdown menu, redundant buttons (go to next season and go to previous season) and slider to select season should be removed. This improvement is added in future work to do. Please see more description of animation in 5.2.2.

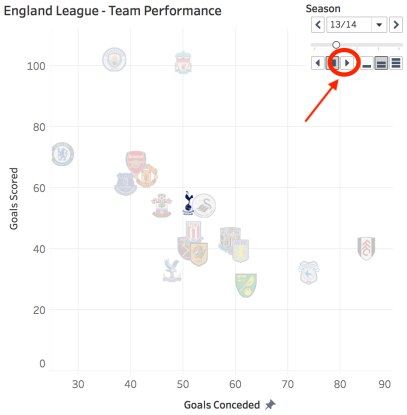

Fig. 21. Play Button for Animation in Task 2

4) message appearing when hovering on help button (Fig. 22)

There is a help button on top right corner for quick guide of interactions for both tasks, in case user is not aware that idioms are interactive. Help message about what interactions are available in each task will be shown on hovering.

Fig. 22. Help Button on Top Right Corner

5 IMPLEMENTATION

5.1 Description

Tableau is used for implementation. The data is stored in Google drive and "Connect to Google Sheets" function is used to read data. Due to lack of human resources in the group, this design is not improved by D3. The suggestions of improving design is given in section 9 (future work).

Visual encoding idioms proposed are relatively easy to implement since they are basic. However, we spent much time implementing interaction idioms and especially the animation for task 2 (see 4.2.6).

5.2 Challenges and Solutions

5.2.1 Button

Tableau does not have button feature, as a consequence, there is no navigation menu bar to locate the specific button for user to navigate menus. In order to implement a menu-like button, a measure of calculated field is created, and then the created field is dragged into the rows. After that, the rows of the 'button' showed label name and a header name together besides the 'button' field. Afterwards, 'hide field label for rows' and then 'hide header' are chosen, and now the calculated field acts as a button.

In addition, actions should be added for navigation buttons. A navigation button is used to send information between worksheets, so that when user clicks it, information of another menu will be shown. However, the most challenging part is the interaction between menus. Since our vis is analyzing four leagues, user needs to click a league from navigation bar to view corresponding vis for that league. This interaction is one-to-many instead of one-to-one. For example, when user clicks German, or Italy, or Spain within England dashboard, the view should be switched to these three leagues correspondingly. That is, within dashboard of a specific league, all other leagues can be selected as destination. So there are many possible actions and it increases the difficulty of implementation.

5.2.2 Animation

For the dashboard 2, animation [20] [22] is added for the scatter plot to demonstrate the change of the team performance. The initial design of animation is different. In the previous version of design, two idioms within task 2 share the same season filter. After user selects a team from stacked bar chart and clicks "show animation" button, the highlighted scatter in scatter plot will move to the position of next season's performance. So that user can easily track the change of performance between two seasons and thus learn the direct effect of the transfer fee.

However, since two idioms share the same filter, Tableau does not support changing data for scatter plot without changing data in stacked bar chart. So the design is modified to use separate filters for two idioms and show animation for seasons afterwards. Please see current animation in demo video.

5.2.3 Pros and Cons

Pros and cons of using Tableau is listed below.

Pros:

1) There are many built-in features, which is convenient to use. For example, showing value when hovering on a badge in scatter plot is a built-in function for scatter plot.
2) There is 'show me' function for idioms, which offers valid idioms for user to choose based on the data provided.
3) The 'drag' feature is super handy, which allows user to drag dimensions, or measures to marks, so that user can edit mark quickly.

Cons:

1) Tableau does not have version control function, so it is hard for a group of people working on the same project.
2) Tableau is not friendly with regard to building dashboard with multiple pages and especially those with hierarchy. There are more than 40 worksheets to finish in order to implement the proposed dashboard. And we can only edit data source during each idiom construction and cannot edit data source within dashboard view. So it takes a lot of time to build idioms, drag them into dashboard, and adjust positions for each idiom within each dashboard. Every time after we update an individual idiom, we have to drag it into dashboard again.
3) The authorization of shared data on Google drive is limited. When the developer shares twb file (the file format of the Tableau), other group members have to log in their own Google account for accessing the data. Even worse, they have to log in for every excel datasheet (more than 20 datasheets) in order to view the entire vis.
4) There is no built-in navigation button in Tableau which limits our implementation. Currently, the navigation effect is simulated by adding actions and filters. Since there are many buttons within navigation bar, many actions and filters are added, which is exhausting and error-prone. It takes a lot of time to debug the navigation effect.
5) For the feature of analyzing data, there is not enough instruction for it, which limits beginners to implement formulas to analyze data. So to implement the stacked bar chart, source data is modified and the transfer fee for one player role of one team (i.e. attack for Chelsea) is summed up using Excel.

6 EVALUATION

6.1 Design Comparison [23]

On a higher level, comparisons are made between different dashboard structures and interaction idioms. For example, which one will be more appropriate for this project, a dashboard with or without tabbing (see 4.2.1)? As for the mid-level, comparisons are made between visual encoding idioms. This method is frequently used during construction of an idiom to evaluate which design element to use. For example, when constructing bar chart for task 1 and 2, comparison between vertical bar chart and horizontal bar chart is made.

6.2 Usability Test

Usability test will be conducted among classmates of this course, and both observation and interview method will be used. Users will be asked to complete all proposed tasks (see 2.2) for English league by themselves since idioms are the same for all leagues. Their actions will be observed and completion time will be recorded. After they complete tasks, they will be asked if they have difficulty with any task, if yes then why and what are possible improvements. We expect to get more concrete and constructive advice for improving idioms from classmates since they have basic visualization knowledge.

6.3 Performance Evaluation

Performance evaluation will be conducted to ensure our design is smooth to use. There are many interaction idioms within our design, so we need to make sure those interactions occur within the reasonable amount of time. Performance will be measured using latency. Latencies are measured for each interaction idiom listed in 4.2.6. Furthermore, tab switching should be smooth, idiom loading should be fast, and a message should appear quickly when user hovers on the bar in bar chart or a scatter in scatter plot. If the latency is longer than 1 second, progress indicator will be added [16].

7 FINDINGS

7.1 Scenarios of Use

Scenarios of use are given for different user tasks (see 2.2). Description and screenshots will be given in detail.

7.1.1 Task 1

For task 1, user Tom wants to examine relationship between money received and team performance for German league, so he clicks “England” and “Money Received & Team Performance” and idioms for task 1 are presented (Fig. 11). He is interested in performance for the most recent season 16/17. Since the default season is 13/14, he uses the filter to select season 16/17 (Fig. 23).

Presented with two idioms (Fig. 24), he is interested which team receives the most amount of money for season 16/17. By checking names of idioms, he knows that the idiom on the left is for money received and the idiom on the right is for team performance. To find the team which receives the most amount of money, he starts to look at the bar chart. That team (Chelsea) is already highlighted using green bar so he can easily find it, and he sees that Chelsea receives 153.2 million £ for that season. And he looks across the bar chart and finds that the team (Sunderland) receives the least amount of money is highlighted by red color.

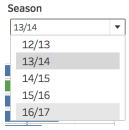

Fig. 23. Filter for Season Selection

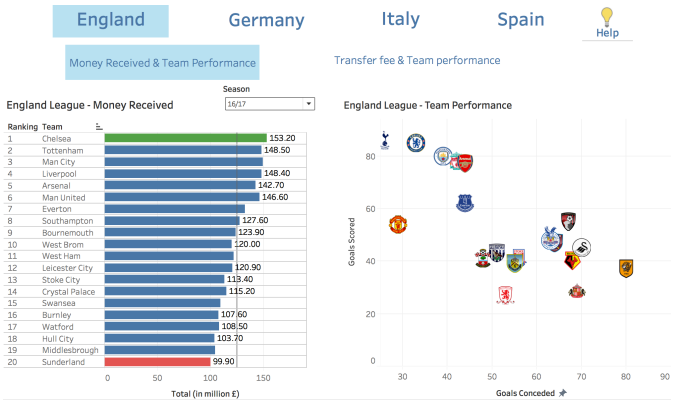

Fig. 24. England Task 1 Season 16/17

He then wants to find out the relationship between money received and team final ranking for one season. And he finds that for season 16/17, the team having the highest ranking receives maximum amount of money and the team having the lowest ranking receives minimum amount of money. Out of curiosity, he wants to check if this is true for other seasons. He checks other seasons for English league, and find that it is somewhat true since sometimes (like season 15/16) the team with the second ranking receives the most amount of money (Fig. 25). And then he checks it for other leagues, he finds that the team having ranking 2 receives the least amount of money for season 16/17 for German league (Fig. 26). So the conjecture is not true.

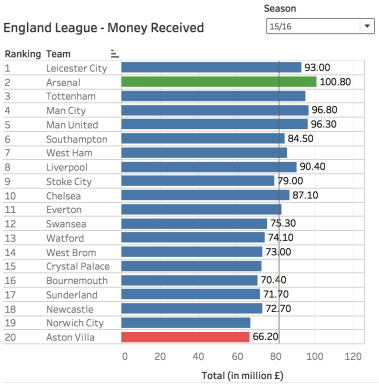

Fig. 25. Bar Chart for England Task 1 Season 15/16

Fig. 26. Bar Chart for Germany Task 1 Season 16/17

Going back to season 16/17 for English league (Fig. 24), he wants to figure out the relationship between team performance and money received. By looking at axes of scatter plot for team performance, he finds that a team with better performance will appear at the top left corner of the scatter plot. And since the scatters in scatter plot are team badges, so he can easily identify teams from

scatter plot. Because Chelsea has the highest ranking for season 16/17, Tom wants to know if Chelsea has the best performance for that season. He sees that Chelsea does not have the best performance (maximal number of goals scored and minimal number of goals conceded), but it has a pretty good performance. Since the actual number of goals scored and conceded are not known, Tom tries to explore scatter plot more by himself.

While he is browsing the dashboard, he finds help button on the top right corner. He hovers on the help button and then knows that the vis is interactive and he can hover on badge of Chelsea to see details. By hovering on the badge of Chelsea in scatter plot, he finds that number of goals scored is 85 and number of goals conceded is 33 for Chelsea in season 16/17 (Fig. 27). And by hovering on a badge in scatter plot, the corresponding bar for that team is highlighted in the bar chart. Tom feels it is really easy to explore relationship between money received and team performance. And he then finds the team with higher ranking tends to have better performance as expected.

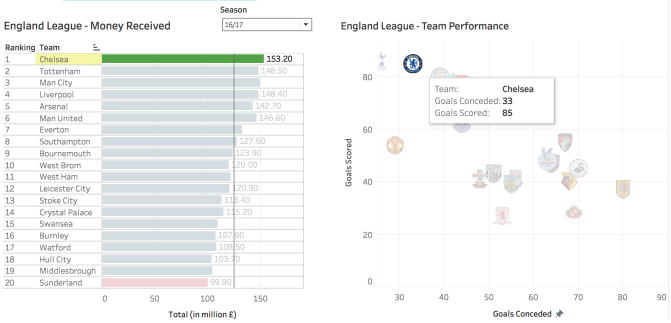

Fig. 27. England Task 1 Season 16/17 – Chelsea Highlighted

7.1.2 Task 2

For task 2, Tom wants to examine relationship between transfer fee and team performance for English league, so he clicks "England" and "Transfer Fee & Team Performance" and idioms for task 2 are presented (Fig. 12). He selects season 13/14 for both idioms and he finds that the stacked bar chart on the left shows money spent on player transfer for teams and the scatter plot on the right is the same as the scatter plot in task 1.

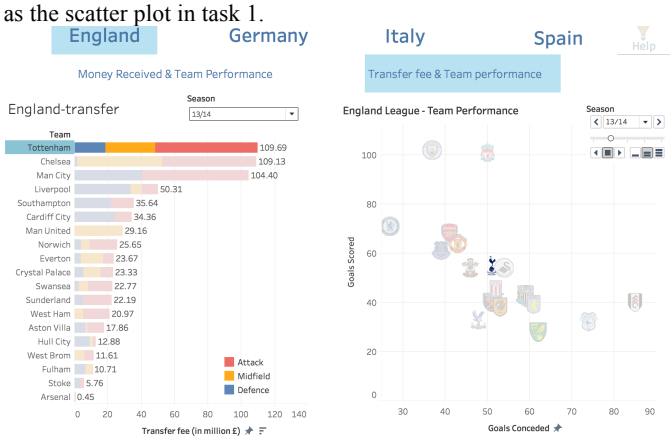

Fig. 28. England Task 1 both Season 13/14 – Tottenham Highlighted

Tom starts to look at the stacked bar chart first and he finds that teams are ordered by the total amount of transfer fee and he finds that Tottenham spends most (109.69 million £) for that season. By looking at the legend he knows that Tottenham spends the most on attack, and more than 50% of transfer fee are spent to bring in attacking players, so he expects number of goals scored of Tottenham to increase for next few seasons. Tom notices that the filter for scatter plot is different from usual filters, so he looks at help menu and learns that he can see animation of performance for one team. So he clicks Tottenham in stacked bar chart (Fig. 28) and finds that the badge for Tottenham in scatter plot is highlighted. After he clicks play button, the badge of Tottenham is moving upwards through time as expected. Since there is only one highlighted moving object, it is pretty easy to trace. Tom is satisfied with the animation.

7.2 System Performance

System performance is tested according to requirements in 6.3 and all interactions respond within 1 second. Interactions triggered by hovering respond even more quickly. So this tool is considered as responsive and smooth to use.

7.3 Feedback

7.3.1 Feedback from Classmates

Feedback gathered from classmates are presented as follows:

1) Add average lines for scatter plot.

Since average line is added for bar chart, we get suggestion that adding average lines for scatter plot. That is, we can add a line to indicate average level of number of goals scored and another line for average number of goals conceded. This suggestion is added into future work section.

2) Why no average line for stacked bar chart?

The user questions that why there is average line for bar chart but not for stacked bar chart. There are several reasons. First, there are different averages to use, averages of total transfer fee or averages of transfer fee for one category. It could confuse user if not enough clarification is given. Second, horizontal stacked bar chart uses vertical lines to segment each portion. Average line, which is also vertical, could interfere with vertical lines within stacked bar chart. The solution could be adding a button to show/hide average line, but it will be visually cluttered when average line is shown.

3) Why not switch axis direction for goals conceded?

Another user suggests that we can switch direction for x-axis (number of goals conceded) for scatter plot. That is, instead of starting from 0 on the left, we can use the maximum of goals conceded on the left, and as the axis goes to the right, number of goals conceded gets smaller. By doing so, the team with worse performance stays close to the origin and the team with better performance will be far away from the origin. This is a nice idea but switching direction of axis is not natural and could be risky. More usability test needs to be done to verify this idea.

7.3.2 Feedback from Colleagues

Feedback gathered from colleagues are presented as follows:

1) Why tabbed dashboard and no support for league-wise comparison?

We are not concerned with comparison between the different leagues for this visualization, and so the issue of cognitive load doesn't arise here since there is nothing to compare. The reason league comparison is not important to us is there are many other factors that are considered when comparing performances of different leagues. For example, if you look at two teams from different leagues, the team teams they play against are completely different. This vis focuses on competition and performance among teams within a league. When different leagues are compared, people look at the how teams from each league perform against each other in the champions league. The champions league is a continental competition that brings most of these teams together and note that it is a different competition. Also, the amount of money that each league receives is different, and the way they divide the money also differs. In summary a comparison of money between teams from different leagues does not offer much insight about the quality of the team or the league.

2) Why no legend for scatter plot?

We get feedback that scatter plot is lack of legend which makes vis confusing and user cannot relate two idioms together.

There are two reasons why legend is not used for scatter plot. First, this vis is designed for soccer fans who have sufficient background knowledge about football and they should be familiar with most team badges. Second, bidirectional interaction is implemented for both tasks. So even if user is not sure which team a badge represents, he can easily see corresponding team name by hovering (task 1) or clicking (task 2) on the badge.

8 Discussion

8.1 Pros and Cons of Design

Since the goal of design is not to optimize but to satisfy, this design has its own benefits and drawbacks.

Pros:

- Dashboard hierarchy is very clear. Separating idioms by task provides only necessary idioms for one task, which reduces confusion and saves pixels for each idiom.
- Bidirectional highlighting is implemented for idioms within both tasks, so user can easily relate idioms together and there is no need to add legend for scatter plot.
- For bar chart, 1) teams are ordered by final rankings so that user can easily detect relationship between money received and final ranking; 2) minimum and maximum bars are highlighted using red and green for easy detection.
- For scatter plot, 1) badges are used for easy distinguishability; 2) axes are adjusted for minimizing occlusion between badges; 3) values will be shown on hovering and can be used for accurate comparison.
- For stacked bar chart, 1) data is aggregated into three categories to achieve a better match with measurements in scatter plot; 2) teams are ordered by total transfer fee descendingly so that min and max can be easily seen.
- Animation is implemented for scatter plot in task 2, so that user can easily track change of performance for one team across years.
- Help button is added in case user is not aware of interaction idioms.

Cons:

- This design aggregated player transfer data so that individual player information (name, team from, transfer fee and detailed position) is not visualized.
- There is no constraint implemented for two filters in task 2, so user can select any year for both filter. It might cause confusion since user is not supposed to select a smaller year for scatter plot.

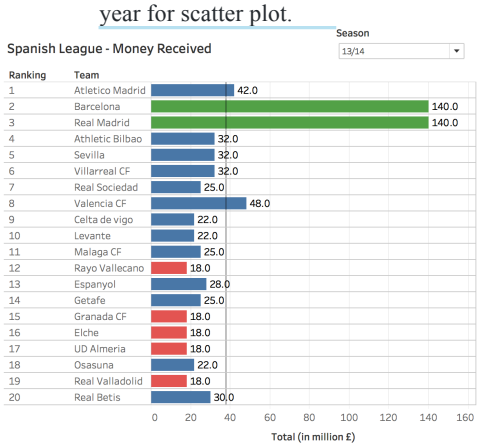

Fig. 29. Spain Task 1 Season 13/14

- Red and green is used to highlight minimum and maximum for bar chart. But if there are several minimum and maximum, it may cause confusion (Fig. 29).

8.2 Lessons Learned

8.2.1 Design as Redesign

Design is a process of validating current design and seeking for other design options. Besides, visualization design is data driven. The most challenging part of this project is to design vis for player transfer data. Initial design is to present data in a table since the data is of high dimension, but this idea is too naive and it simply presents raw data. Later, the idiom is redesigned as chord diagram. But chord diagram is not applicable for our data and it is challenging to implement using Tableau. The final design uses stacked bar chart since it matches well with measurements in scatter plot (see 4.2.5). Since all visual encoding idioms used are relatively simple, we can exert effort on implementation of interaction idioms.

8.2.2 Design and Implementation as a Whole

During design, the designer should consider difficulty of implementation. Besides, the designer should talk to the developer often and update design as needed. Though some design idea is valid, it could be challenging to implement. If so, other design options should be considered.

When designing animation for task 2, initial plan is to use one season filter since the transfer fee spend for this season will have direct impact on performance on the next season. And if "show effect of player transfer" button is clicked, animation will be shown for mark movement (for a selected team) in scatter plot from current season to the next season. However, developer tried and cannot find a way to implement it with Tableau. So animation is redesigned as the current version.

8.2.3 Data Preprocessing is Crucial

Data preprocessing should be started as early as possible and we cannot assume the data is ready to use. Before starting implementation, we thought we have all the data and data preparation is done. However, the data collected needs additional cleaning. For example, team names we have are not unified, "AC Milan" and "Milan" are the same thing but they both occurred within raw data. Some money spent field has only numbers, some others contain character "m" in it indicating million. It took us a significant amount of time to preprocess data. Without data preprocessing, we may misinterpret data or give different conclusions via visualization.

9 Future Work

9.1 Improve Design using D3

As Tableau has its drawback on button creation (see section 5) and interactions created are constrained, the design proposed can be implemented and improved using D3 in the future. Compared to the current navigation bar (Fig. 30), we can build a tabbed navigation bar (Fig. 31) using D3 [24]. Besides, we will be able to implement the animation mentioned in 8.2.2.

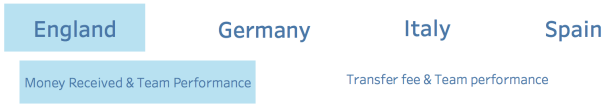

Fig. 30. Current Navigation Bar

Fig. 31. Desired Navigation Bar

There is no built in button in Tableau, so it is time-consuming to implement buttons. But for D3, it is relatively easy to create buttons and then we can put a "Show detail" button beside stacked bar chart. When user clicks that button, stacked bar chart will be replaced by a table showing detail of player transfers, and user can click "Go back to overview" to go back to stacked bar chart view.

9.2 Add Word Cloud for Transfer Data

Word cloud (Fig. 32) can be used to visualize overview of new players arriving. This idea is inspired by "football subreddit word cloud [25]". There will be one word cloud generated for each team during a season. The size of string encodes individual player transfer fee, larger string means higher transfer fee; the color of string encodes different players. Since there are not too many players arriving at one team during a season, number of bins for color is sufficient [16]. Individual player transfer fee and team from will be shown when hovering on a player name in word cloud. Note that there are players bought for free, in this case, minimum text size should be set and legibility of text should be considered.

Fig. 32. Player Transfer Vis - Word Cloud

9.3 Miscellaneous

- For scatter plot, add average lines to indicate average number of goals scored and conceded (see 4.2.4 for scatter plot).
- For stacked bar chart, change order of colors within legend to make it match with colors within one bar (see 4.2.5 for stacked bar chart).
- The first time user opens this vis, help button should blink to attract user's attention and guide user about how to use this vis (see 4.2.6 for more detail on help button).
- For animation in task 2, constraint should be added so that the user cannot play animation without selecting a team.
- For animation menu in task 2, remove redundant "go to previous" and "go to next" buttons and slider for season selection (see 4.2.6 for more detail on animation menu).

REFERENCES

[1] Twitter. (2018). Premier League: Where are your club's followers?. [online] Available at: https://interactive.twitter.com/premierleague/#?mode=team&teamId=all

[2] Ou, J. (2018). *Insights in European league soccer players transfer*. [online] Eng.utah.edu. Available at: http://www.eng.utah.edu/~jianil/process_book/process_book.html

[3] Jimbotemps.github.io. (2018). Premier League Wage Data Visualisation. [online] Available at: https://jimbotemps.github.io/

[4] Widget Store. (2018). Soccer Average Goals Scored & Conceded Widget Sportradar Widget Store. [online] Available at: https://store.sportradar.com/en/widgets/soccer/soccer-average-goalsscored-conceded.php

[5] Ohyoubeauty.blogspot.ca. (2018). *Liverpool Goals Scored and Conceded 2016-17*. [online] Available at: https://ohyoubeauty.blogspot.ca/2017/05/liverpool-goals-scored-and-conceded.html [Accessed 10 Apr. 2018].

[6] Espnfc.com. (2018). *Transfers - ESPN FC*. [online] Available at: http://www.espnfc.com/transfers [Accessed 10 Apr. 2018].

[7] euros), S. (2018). *Soccer players highest transfer fees all-time 2018 | Statistic*. [online] Statista. Available at: https://www.statista.com/statistics/263304/transfer-fees-the-10-most-expensive-transfers-in-soccer-ever/ [Accessed 10 Apr. 2018].

[8] Keating, J. (2018). Fun Map Shows 113 Years of Soccer Player Migration Around the World. [online] Slate Magazine. Available at: http://www.slate.com/articles/sports/foreigners/2013/11/football_migration_113_years_of_soccer_player_transfers_on_one_map.html

[9] reddit. (2018). Visualization of Jorge Mendes's player transfers between the same teams. • r/soccer. [online] Available at: https://www.reddit.com/r/soccer/comments/5ru5cl/visualization_of_jorge_mendess_player_transfers/

[10] Cs.ubc.ca. (2018). [online] Available at: https://www.cs.ubc.ca/~tmm/courses/547-17/projects/yann/report.pdf [Accessed 10 Apr. 2018].

[11] Tableau Software. (2018). *Data Story: The History of Football Player Transfers*. [online] Available at: https://www.tableau.com/solutions/gallery/history-football-player-transfers [Accessed 10 Apr. 2018].

[12] Premierleague.com. (2018). Premier League Table, Form Guide & Season Archives. [online] Available at: https://www.premierleague.com/tables

[13] Footballdatabase.com. (2018). FootballDatabase - Club Rankings and Statistics. [online] Available at: http://footballdatabase.com

[14] Transfermarkt.co.uk. (2018). Football transfers, rumours, market values, news and statistics. [online] Available at: https://www.transfermarkt.co.uk

[15] BI Blog | Data Visualization & Analytics Blog | datapine. (2018). 10 Dashboard Design Principles & Best Practices To Convey Your Data. [online] Available at: https://www.datapine.com/blog/dashboarddesign-principles-and-best-practices/

[16] Munzner, T. (n.d.). Visualization analysis & design.

[17] Choudhury, S. (2018). Choosing the right chart type: Bar charts vs Column charts - FusionBrew. [online] FusionBrew - The FusionCharts Blog. Available at: https://www.fusioncharts.com/blog/bar-charts-orcolumn-charts/

[18] YouTube. (2018). *How To Make Hover Text Help in Tableau - Datasaurus-Rex*. [online] Available at: https://www.youtube.com/watch?v=Es7aSLs4qKI [Accessed 10 Apr. 2018].

[19] Inviso.dk. (2018). *Labels on stacked bar chart | Inviso*. [online] Available at: https://inviso.dk/blog/post/labels-on-stacked-bar-chart [Accessed 10 Apr. 2018].

[20] YouTube. (2018). *Tableau animation in five minutes*. [online] Available at: https://www.youtube.com/watch?v=b9AWjOtWQSM [Accessed 10 Apr. 2018].

[21] Csc2.ncsu.edu. (2018). *Perception in Visualization*. [online] Available at: https://www.csc2.ncsu.edu/faculty/healey/PP/index.html [Accessed 12 Apr. 2018].

[22] YouTube. (2018). *Creating animated Tableau dashboards | Pluralsight*. [online] Available at: https://www.youtube.com/watch?v=iLkSvKqbp1c [Accessed 10 Apr. 2018].

[23] Innovis.cpsc.ucalgary.ca. (2018). [online] Available at: http://innovis.cpsc.ucalgary.ca/innovis/uploads/Publications/Publications/Lam_2011_SGS.pdf

[24] Cssmenumaker.com. (2018). *Flat Tabbed Menu | CSS MenuMaker*. [online] Available at: http://cssmenumaker.com/menu/flat-tabbed-menu [Accessed 10 Apr. 2018].

[25] Kvsingh.github.io. (2018). *Football subreddit word clouds - Visualizing transfer window madness // UnInhibited Zombie //*. [online] Available at: https://kvsingh.github.io/soccer-subreddit-wordcloud.html [Accessed 10 Apr. 2018].

Appendix A: Video and Slides

Demo video and slides for final presentation are pushed to project-team--alpha-go repository (private) on GitHub. The video and slides can also be accessed via link (please copy and paste into browser): https://drive.google.com/drive/folders/1F-uDnXoTxYqgMR64MWIboIuce-j0YXIf?usp=sharing

Appendix B: Milestones

Table 1. Project Milestones

What to Do	Who	When to Deliver
Interim Product:		
Data Preparation	Mohamed	Feb 20
Vis Design	All	Feb 24
Vis Implementation	Rui	Mar 4
Interim Report (due Mar 5):		
Introduction and Motivation	Mohamed	Feb 24
Related Work	Mohamed, Shiyi	Feb 24
Approach and Milestones	Shiyi	Feb 24
Implementation and Technology Choice	Shiyi	Mar 2
Proposed Evaluation Steps	Shiyi	Feb 27
Risks	Mohamed, Rui	Mar 2
Document Formatting	Shiyi	Mar 4
Final Product:		
Data Cleaning	Rui	Mar 20
Idiom Improvement	Shiyi	Mar 21
Finalized Implementation	Rui	Mar 29
Usability Test Conduction	Rui	Apr 6
Demo Video (due Apr 5):		
Video Script	Shiyi	Apr 1
Voice-over	Mohamed, Rui	Apr 3
Final Presentation		
Slides Preparation	Shiyi, Mohamed	Apr 3
Presenting	Mohamed	Apr 5
Final Report (due Apr 12):		
Introduction	Mohamed	Apr 2
Implementation	Rui	Apr 8
Other sections	Shiyi	Apr 10
Document Formatting	Shiyi	Apr 11

Appendix C: Usability Test - Tasks

Feedback from usability test for tasks is already included in section 7.3.1, and the design of usability test for performing tasks is given as below.

Ask user to perform following tasks for English league without additional instruction. See what they will do, and if they will find help button on the top right. For each task, record if it is accomplished successfully or not, record time to perform. If a task is not completed successfully, ask user why or what is confusing. Ask user if he will notice the vis is interactive by himself. Note that tasks are defined at the beginning of report in section 2.2. Besides, task 2.2 is not tested since it is the same as task 1.3.

Note that since user feedback is long, it is listed in a separate table.

Table 2. Usability Test – Task Completion

Task	Completed?	Completion Time (if completed)
1.1	Yes	4s
1.2	Yes	4s
1.3	Kind of	1min 5s
1	Yes	18s
2.1	Yes	25s
2	Yes	1min 23s

Table 3. Usability Test – User Feedback

Task	User Feedback
1.1	User says that he did not focus on the values provided. His first catch is the 1st position instead of the green color, which is a little bit confused.
1.2	/
1.3	He struggled for a while, since he has a little football knowledge but not much. He thought the performance of a football team should be number of goals scored minus number of goals conceded. He understands it after I explained to him about the metrics we use.
1	He questions our design about the scatter plot since there is no instructions for the scatter plot. And he mentioned that we did not define which kind of performance should be regarded as 'good' or 'bad' within dashboard. He suggests us to swap the direction of the goal conceded (x axis) from 0 - 90 to 90 - 0. He said since the very bottom of the Y axis means the lowest goal scored. Based on this case, the very left point of the X axis should be the worst case too. Then our entire scatter plot will be like this:  That is, the farther and the higher badge is, the better team performance is.
2.1	/
2	He suggests us to swap the x axis as well for this case.

Appendix D: Usability Test - Performance

Test performance of application, see if latency is less than 1 second. If yes, that's good. Record it if latency is larger than 1 second. Test average response time for elements listed in the table below, ask if user feels the interaction is smooth.

Table 4. Usability Test - Performance

Element Tested	Smooth?	User Feedback
Help button	Yes	It's useful.
interaction between idioms (hover bar in bar chart, hover on scatter in scatter plot)	Yes	/

interaction between idioms (click team name in stacked bar chart, click scatter in scatter plot)	Yes	User asks why it uses hovering for task 1 but clicking for task 2. This question is answered in section 4.2.6.
animation after play button is pressed	Yes	User is confused with the slider within animation playing menu. Clarification for animation menu is in section 4.2.6 and improvements to animation is added into future work.