

Know your audience

DATA VISUALIZATION IN POWER BI



Dashboards versus "pixel-perfect" reports

Dashboard

- Tells us about critical business or operational metrics, often in real time
- "Single pane of glass"
- The norm for Power BI (even when we call them reports!)

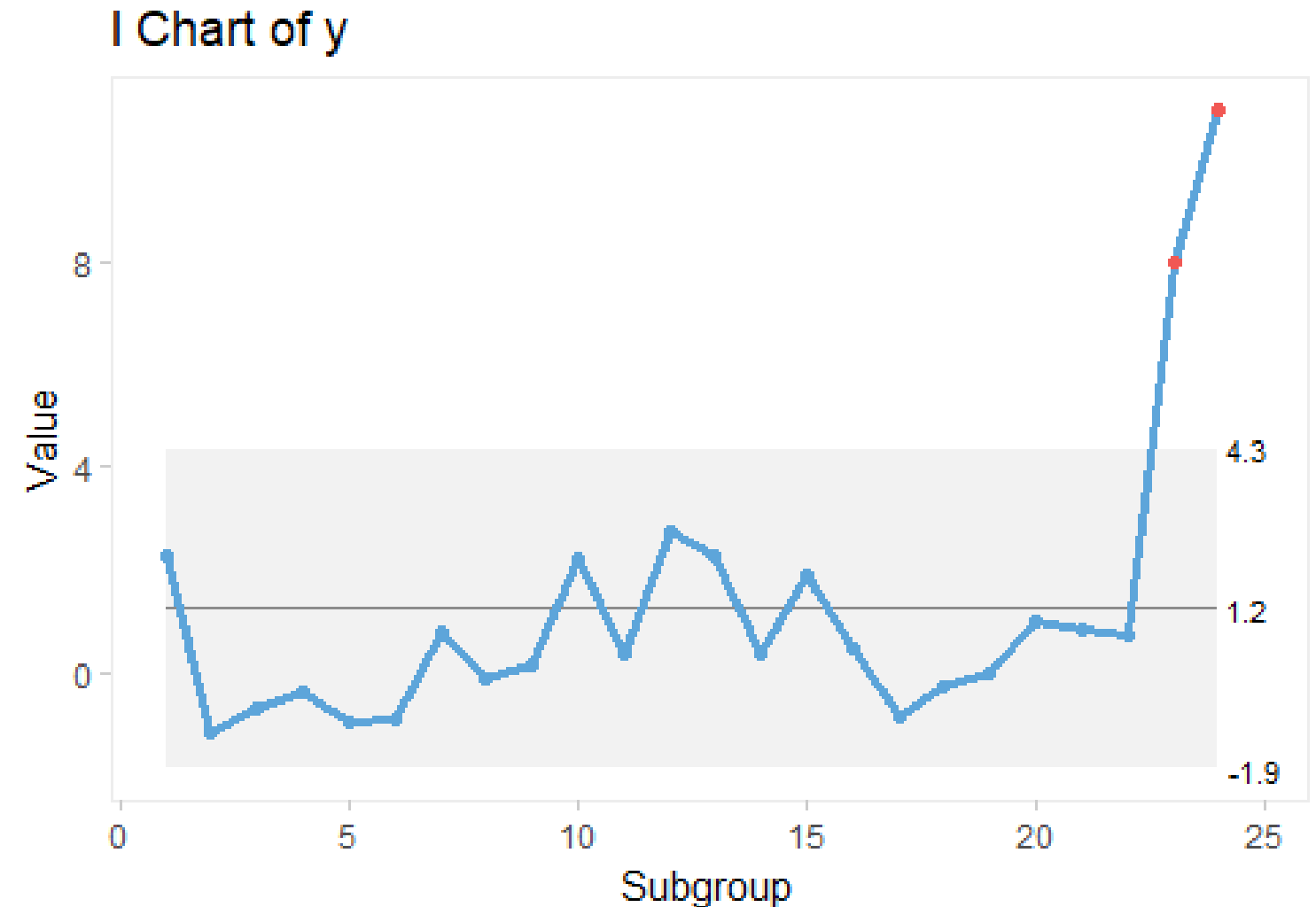
"Pixel-Perfect" Reports

- Provides more detailed information on business or operational metrics, usually delayed
- Intended to print on sheets of paper, not just on screen
- Called "Paginated Reports" in Power BI

Your intended audience

Questions to ask

- Who is your intended audience?
- How will they use your dashboard?
- What actions do you want them to take as a result of what you show?
- Are you showing the right measures in the right way?



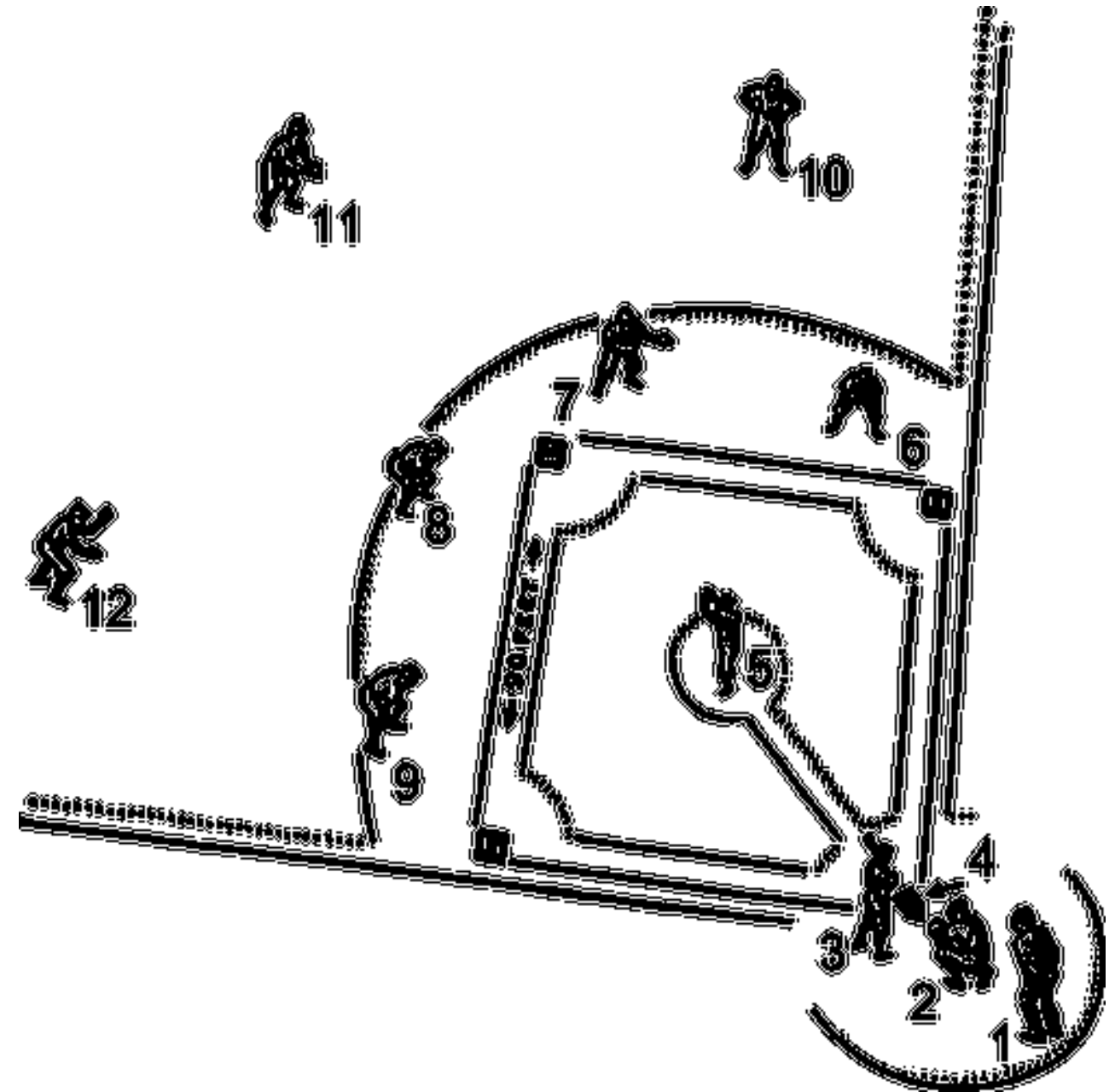
The dataset

Lahman's Baseball Database

- Complete batting and pitching statistics from 1871 to 2020
- Includes fielding statistics, standings, team statistics, etc.
- Available in several formats (including an R package!)

A primer on baseball

- Played on a diamond with home plate (4) and three bases.
- Games typically last 9 innings. Each team bats (that is, plays offense) until they commit 3 outs.
- 9 players on the field for the defense, including the pitcher (5) and catcher (2).
- One person bats at a time (3). Each base may have 0 or 1 baserunners.
- Goal: get on base. Then, score by touching home plate after touching all three bases in



¹ <https://publicdomainvectors.org/en/free-clipart/Baseball-diagram/62196.html>

Batting measures and statistics

- At-Bat (AB)
- Hit (H)
- Batting Average (AVG) = $\frac{H}{AB}$
- Walk (BB)
- Hit by Pitch (HBP)
- Sacrifice Fly (SF)
- On-Base Percentage (OBP) = $\frac{H+BB+HBP}{AB+BB+HBP+SF}$



Our audience

The Baseball Historian

- Has knowledge of the sport
- Wishes to dispel or confirm "folk knowledge"
- Does not work for a franchise (team)

Dashboard or Report?

Dashboard or report?

Keep your audience's intent in mind when determining whether you should create a dashboard or a "pixel-perfect" report. Given the following scenarios, when would you recommend a dashboard and when would a report be better?

Our regulatory compliance officer needs a print-out of every unauthorized access attempt for a given month.

The CEO wants to track sales measures and the company's stock price over the course of the day.

A machine worker needs to monitor temperatures at real time.

The Accounts Payable department needs a list of all outstanding invoices we need to pay.

Let's practice!

DATA VISUALIZATION IN POWER BI

Tables and scatter charts

DATA VISUALIZATION IN POWER BI



Let's practice!

DATA VISUALIZATION IN POWER BI

Add a Table and a Slicer Exercise

In this first exercise, we will add a table with some basic information about hitting: number of at-bats, number of hits, and batting average. We will also add a pair of filters for year and franchise, allowing for more granular analysis for our baseball historian persona.

Step 1

Add a new *Table* visual to the canvas and move it to the top-left corner. On this new visual, add the following columns from the `Batting` table: player name and team (`nameAndTeam`), number of at-bats (`AB`), number of hits (`H`), and batting average (`AVG`). Stretch the table down to the bottom of the canvas and approximately 1/3 of the way horizontally. This should be enough to ensure that all columns are visible.

For a better understanding of the dataset you can check the [metadata sheet](#).

Step 2

Add a slicer with `Batting[yearID]` as the *Field*. This refers to the year (`yearID`) on the `Batting` table and is the syntax we will use throughout the course.

Step 3

Add another slicer with `TeamsFranchises[franchID]` as the *Field*. Fit these two slicers on the right-hand side of the visual.

What was the batting average of the player on franchise "BRG" with the largest number of at-bats?

Create a Scatter Plot Exercise

During the 1920s, Major League Baseball introduced rule changes which resulted in the "live ball" era, making it easier to hit the baseball for power. Prior to that, hitting a home run was a fairly rare occurrence. This live ball era also affected the number of doubles, as that is another good indicator of power. Our baseball historian knows that in the modern game, power hitters tend to hit a lot of doubles as well as home runs, and would like to understand whether the relationship between doubles and home runs has remained consistent over time. We will use a scatter plot (or scatter chart) to figure this out.

Step 1

Ensure that the franchise "BRG" has been de-selected and sort the table by `nameAndTeam` ascending.

Step 2

Add a new scatter plot to the page, taking up the upper half of the remaining canvas. Add number of doubles (`Batting[2B]`) to the *X Axis* and number of home runs (`Batting[HR]`) to the *Y Axis*. Add `Batting[yearID]` to the *Values* field.

Step 3

Narrow down the year range from 1920-2020.

What is the greatest number of doubles hit in a single season?

Bubble Chart Exercise

A bubble chart is a scatter chart where the size of the bubble represents a measure. This allows us to track three separate measures on a single visual. In this case, we will track the relationship between on-base percentage (OBP) and batting average (AVG) for players, using the number of at-bats (AB) as the size of the chart. On-base percentage is an "extension" of batting average: with batting average, we track how many hits you had over at-bats. With on-base percentage, we add things like walks and hits by pitch to both the numerator and denominator. Therefore, AVG and OBP are very highly correlated.

Step 1

Add a scatter plot to the page, taking up the remaining canvas. To make this a bubble chart, you will use three measures. The *X Axis* will contain batting average (`AVG`), the *Y Axis* will contain on-base percentage (`OBP`), and the *Size* will contain at-bats (`AB`). Add `nameAndTeam` to the *Values* field.

Step 2

Use the slicers to filter down to the year 1995 and the franchise "ATL" to view details on the 1995 Atlanta Braves.

Step 3

Add a filter on the bubble chart to Include only players with at least 100 at-bats.

How many at-bats did the player with the highest on-base percentage (OBP) have for the 1995 Atlanta Braves?

Challenge

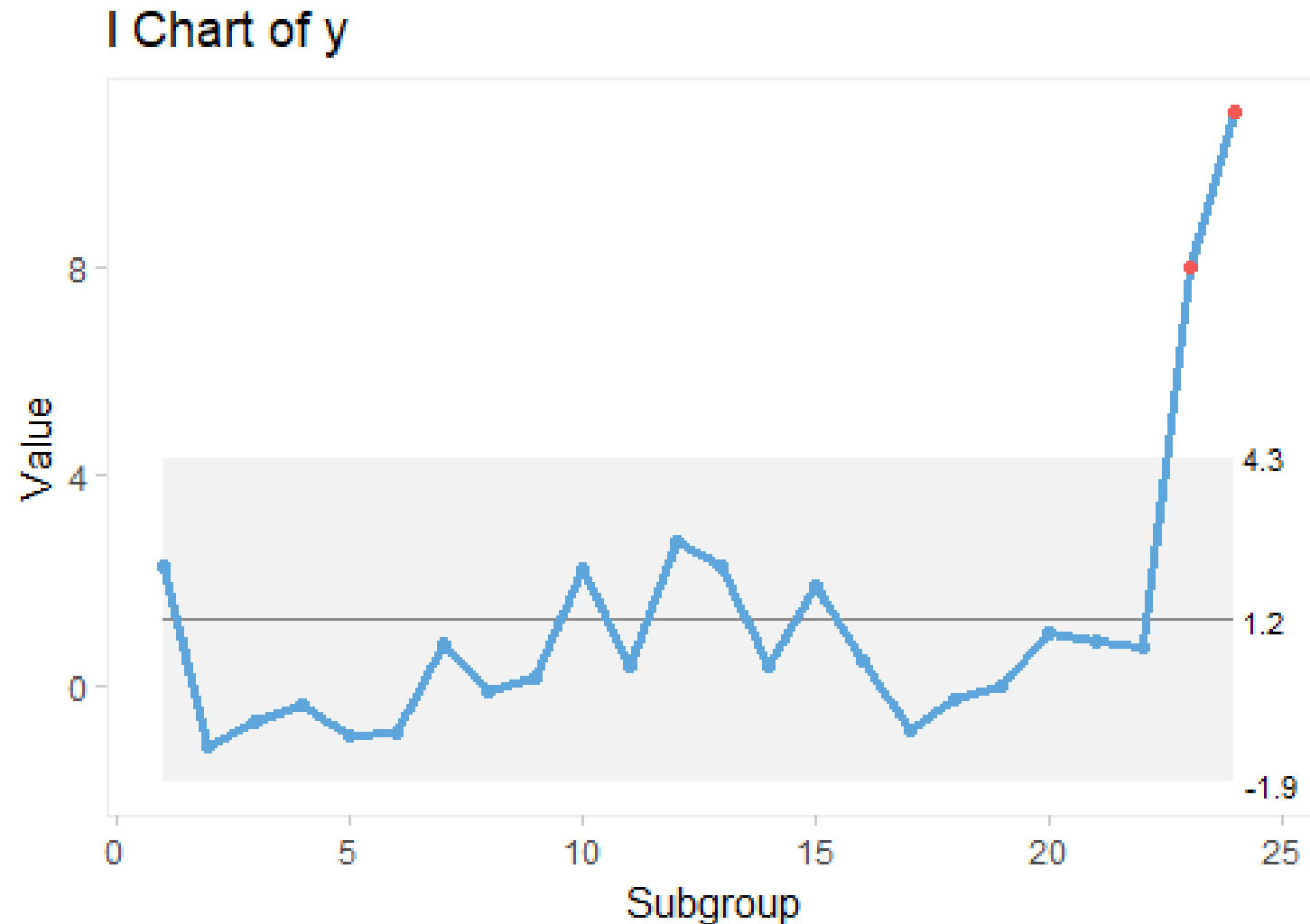
1. Identify 5 INSIGHTS from the Miami Marlins and display them as a PowerBI Visual.
2. Create a Dataset of NBA data and identify 5 INSIGHTS.

Getting an emotional response

DATA VISUALIZATION IN POWER BI



Have you hit the right audience?



Context

- Show the bore depth on drilled machinery versus ideal
- If this is outside the acceptable (grey) zone, we have permanently damaged the product
- You are a specialist responsible for managing this drill press

Emotion engagement

Why I'm Happy



- Everything on the dashboard looks fine
- I do not need to perform any action (or can perform a positive action)



Why I'm Angry



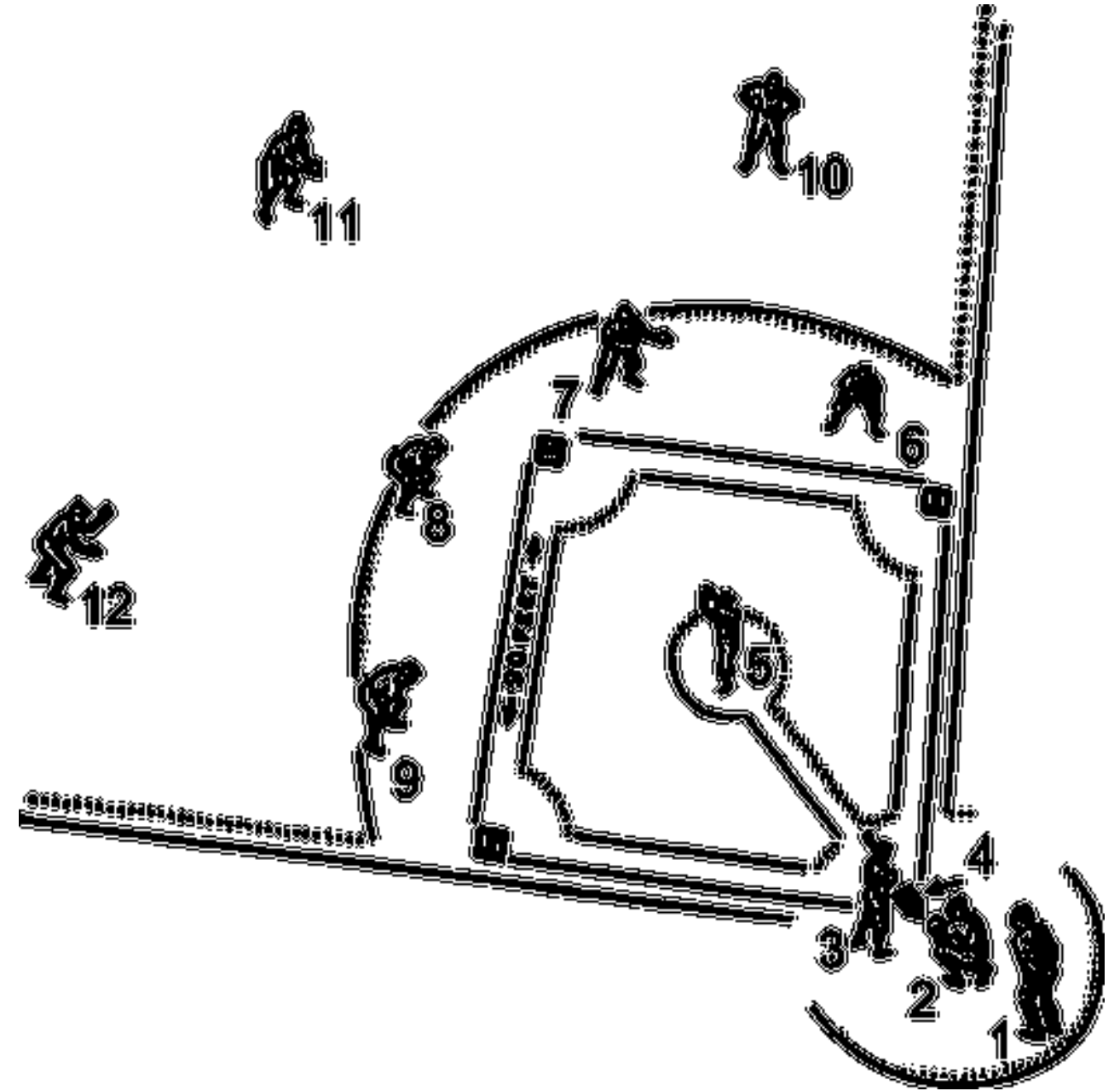
- Something has gone wrong
- Now I need to act in response to this issue

Worst case: zero emotional response

Continuing our baseball primer

Defensive Positions

- Pitcher: 5 (P)
- Catcher: 2 (C)
- First Baseman: 6 (1B)
- Second Baseman: 7 (2B)
- Third Baseman: 9 (3B)
- Shortstop: 8 (SS)
- 3 Out elders: 10, 11, 12 (OF)



¹ publicdomainvectors.org/en/free-clipart/Baseball-diagram/62196.html

Making an out

- Putout
- Assist
- Double play (and triple play)
- Error
- Outs played



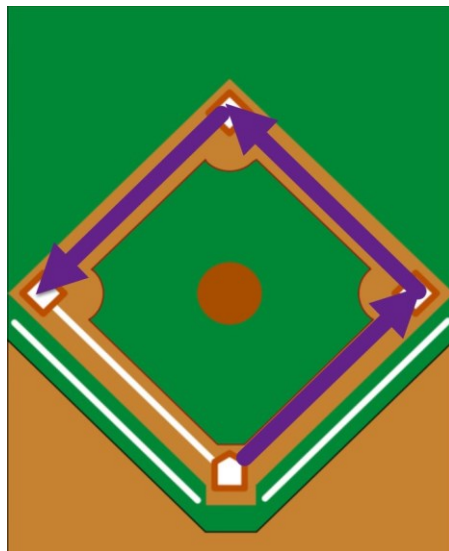
¹ Photo by Ben Hershey on Unsplash

Hitting for power

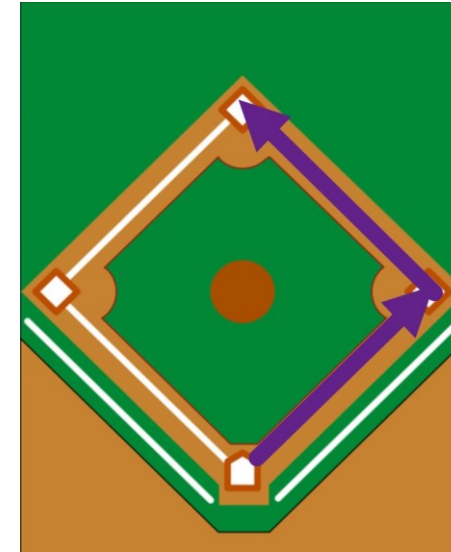
Single (1B)



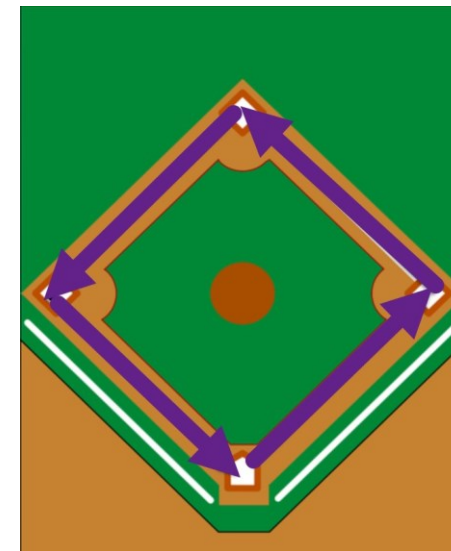
Triple (3B)



Double (2B)



Home Run (HR)



Calculating slugging percentage (SLG)

- Slugging percentage (SLG) = $\frac{\sum \text{Bases}}{\text{At-Bats}}$
- $\text{SLG} = (1\text{B} + 2*2\text{B} + 3*3\text{B} + 4*\text{HR}) / \text{AB}$
- $\text{H} = 1\text{B} + 2\text{B} + 3\text{B} + 4\text{B}$
- $\text{SLG} = (\text{H} + 2\text{B} + 2*3\text{B} + 3*\text{HR}) / \text{AB}$
- Single (1B) = 1 base
- Double (2B) = 2 bases
- Triple (3B) = 3 bases
- Home Run (HR) = 4 bases

Our audience

General Manager (GM)

- Signs players to contracts
- Hires team managers
- Oversees "back office" activities

Let's practice!

DATA VISUALIZATION IN POWER BI

What Evokes an EMOTIONAL Response?

Visuals include information
relevant to the viewer

Visuals have the correct
supporting context

Visuals provide a clear
signal of which action the
user should take

Visuals are accessible to
users with color vision
deficiency

Visuals have clear titles
and use sufficiently large
font sizes

Bar charts and small multiples

DATA VISUALIZATION IN POWER BI



Let's practice!

DATA VISUALIZATION IN POWER BI

Create a Bar Chart Exercise

The general manager (GM) of a team needs to know how our team's players stack up versus the rest of the league. Our first assignment as assistant to the GM is to review the list of starting shortstops in the last full season (2019). The GM has expressed an interest in finding who the best starters in the league were, both to compare against what our team has on the field as well as to open up potential trade negotiations. To do this, we will start with a simple comparison of players by batting average.

Step 1

Add a new report page and name it
Player Comparison .

Step 2

Add a new *clustered bar chart* visual. It should take up the entire vertical area of the canvas and approximately 70% of the horizontal space, starting from the left edge. Create your chart with the batting average by player name + team (that is, `nameAndTeam`) using the `Batting` table.

Step 3

Add a slicer and filter by position (`Fielding[POS]`).
Narrow the list down to shortstops ("SS").

Step 4

Add a page-level filter where the year is 2019. Use the year column from the `Batting` table for this.

Step 5

Add a slicer and filter by outs played (`Fielding[innOuts]`). Narrow the range from 1000 to the maximum available, to capture players who spent a significant time on the field.

Step 6

Add a slicer and filter by at-bats (`Batting[AB]`). Narrow the range from 150 to the maximum available.

What is the name and team of the shortstop with the fourth-highest batting average in 2019 among qualifying players? Answer in the form "First Last, TEAM" as it is displayed in the visual.

Format the Visual Exercise

The Mendoza Line is named after Marlo Mendoza, a light-hitting shortstop who played during the 1970s. His batting average always fluctuated around 0.200, and thus the Mendoza Line is 0.200 and represents a floor for batting averages on baseball players. In the modern era, we might want to adjust that line up to .250.

Also, our GM might be interested in seeing the work we've done so far, but we want to make sure the page looks nice for him.

Step 1

Change the title on the bar chart visual from "AVG by nameAndTeam" to "Batting Average Leaders".

Step 2

Change the display of batting average on the bar chart to display batting averages with 3 digits after the decimal.

Step 3

Add a constant line at 0.250. To see this line better, make sure that the color is "Black, 0% darker" and transparency is 50%.

In baseball, anyone batting below .200 is considered below the Mendoza line, making them an incompetent hitter. In this modern interpretation, threshold is set to 0.250.

Step 4

In addition to this modern Mendoza Line, It would be nice to add a line representing the average batting average. Add this line and set its color to "Black, 20% lighter" with 50% transparency to differentiate it from the modern Mendoza Line.

Step 5

Rename the slicers from "POS" to "Position", "InnOuts" to "Outs Played", and "AB" to "At-Bats".

What is the name and team of the shortstop who hit closest to our modern Mendoza Line in 2019? (Format: First Last, TEAM)

Stacked Errors by Position

Our general manager likes our player comparison analysis and now wants to get a feeling for team defenses, as we have two excellent defensive prospects who we might be willing to trade. To find a partner, we want to figure out which teams had the most difficulty fielding the ball in 2019. Here, we can use a stacked bar chart to get an impression of how many errors each team committed, and break it down by position for further analysis.

Step 1

Add a new page and name it
Team Fielding by Position .

Step 2

Add a page-level filter where the year is 2019.

Step 3

Add a *stacked bar chart* visual. Chart the number of errors (`Fielding[E]`) by franchise (`TeamsFranchises[franchID]`), broken out by fielding position (`Fielding[POS]`).

Step 4

Change the chart title to "Errors per Position by Team".
Hide the *Y axis* title. Change the *X axis* title to "Number of errors".

**Which position group committed the most errors for the Detroit Tigers (franchise ID "DET") in the year 2019?
Enter the answer exactly as shown in the legend.**

Small Multiples Exercise

Small multiples are a way of showing comparison information across two categorical dimensions, saving a bit of space and a lot of time over making individual charts for each combination of factors. In this exercise, we will look at National League (NL) teams in the year 2019 and compare team-wide batting averages at each player position for our GM.

Step 1

Create a new page and call it

Batting Average by Team by Position . Add page-level filters where the year (`Batting[yearID]`) is 2019, the league (`Batting[lgID]`) is NL, and fielding position (`Fielding[POS]`) is not P. That is, we want to avoid including pitchers in this list, as pitchers normally tend to be terrible batters.

Step 2

Add a new clustered column chart visual and have it fill the entire canvas, mapping batting average (`Batting[AVG]`) by franchise (In the `TeamsFranchises` table).

Step 3

Add fielding position to the *Small multiples* field. Then, format the visual and ensure that the grid layout is 3 rows by 2 columns.

Pittsburgh (PIT) was the NL franchise with the highest batting average for first basemen (1B) in the year 2019. What was that batting average?