

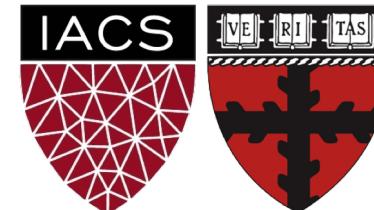
Lecture 14: Visualization

A vehicle for EDA and communication

Harvard IACS

CS109A

Pavlos Protopapas, Kevin Rader, and Chris Tanner



As the matplotlib thickens ...

ANNOUNCEMENTS

- Homework 3 is due Oct 7 (Wed) @ 11:59pm
- Didn't submit **Milestone 1**? Want to change project selection?
E-mail the helpline at cs109a2020@gmail.com
- After lecture, please update your Zoom to the latest version (5.3.1).

Learning Objectives

- Understand why visualization/plotting is important
- Learn aspects that tend to make visualizes **effective** and **ineffective**
- Feel comfortable designing plots that best convey your message
- Gain experience in producing plots with Python



be more cognizant of broader design choices
(e.g., typography, spacing, colors)

Agenda

- EDA Refresher
- Effective Visualization
 - Graphical Integrity
 - Scope
 - Displays
 - Sensible Design
- Communication
 - Motivation
 - Key Considerations

Agenda

- EDA Refresher
- Effective Visualization
 - Graphical Integrity
 - Scope
 - Displays
 - Sensible Design
- Communication
 - Motivation
 - Key Considerations

Assume you know a given dataset is credible,
complete with the info you want, and has no
missing values.

Why do further EDA?

Purposes of EDA:

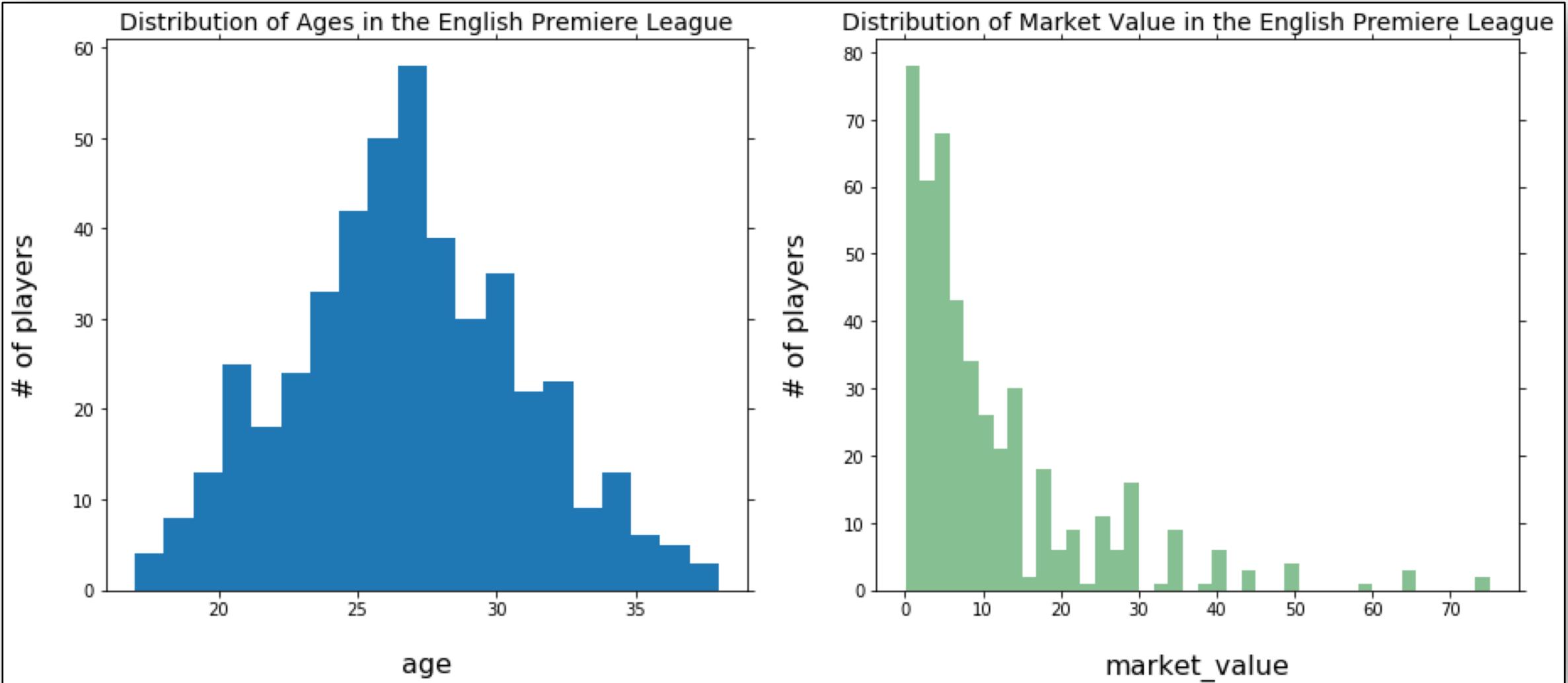
- Maximize insight into a dataset
- Uncover underlying structure
- Detect outliers
- Test underlying assumptions
- Develop parsimonious models

EDA Refresher: English Premier League

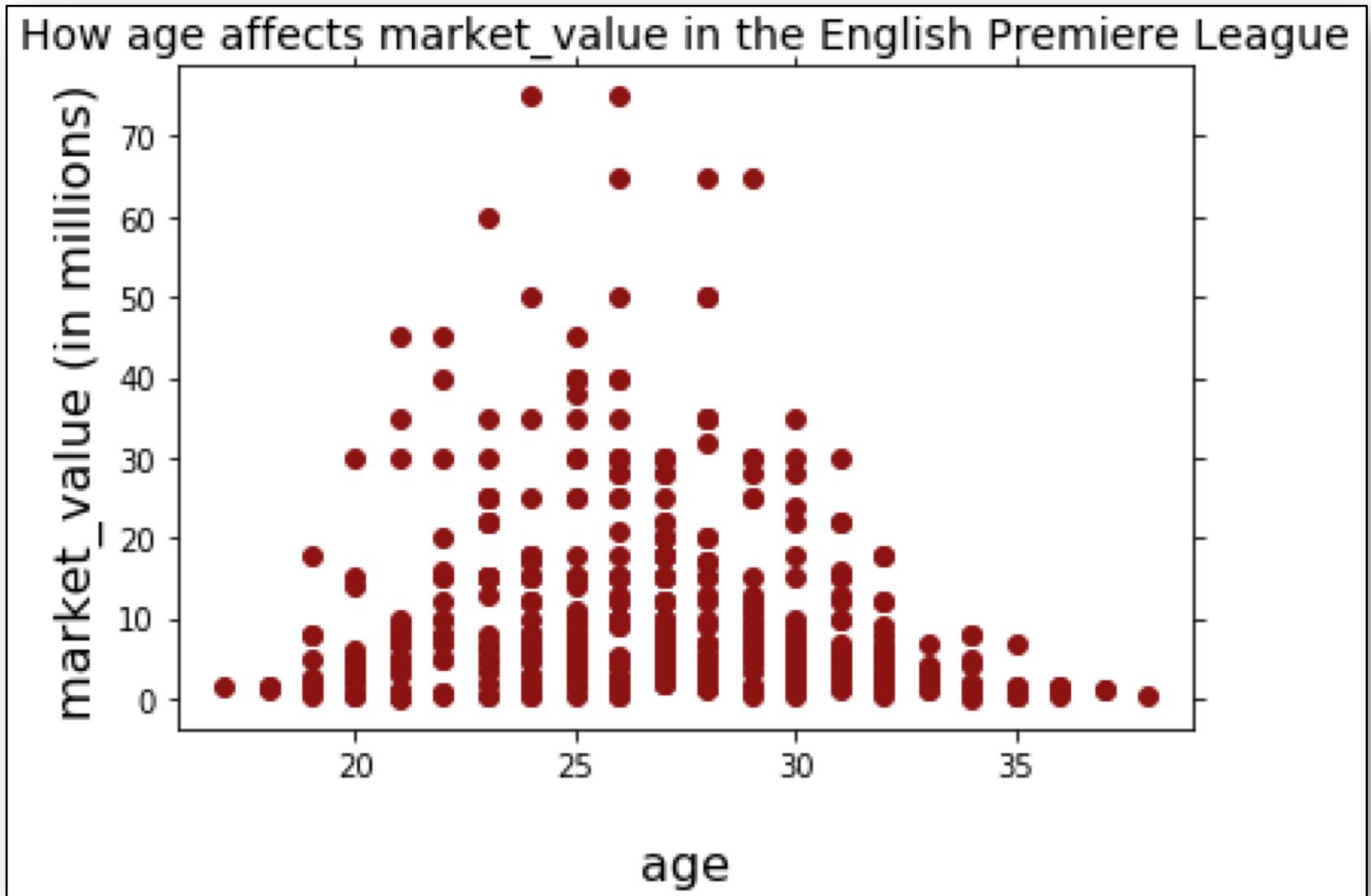
name	club	age	position	market value
Alexis Sanchez	Arsenal	28	LW	65
Mesut Ozil	Arsenal	28	AM	50
Petr Cech	Arsenal	35	GK	7
Theo Walcott	Arsenal	28	RW	20
Laurent Koscielny	Arsenal	31	CB	22

from www.transfermarkt.us

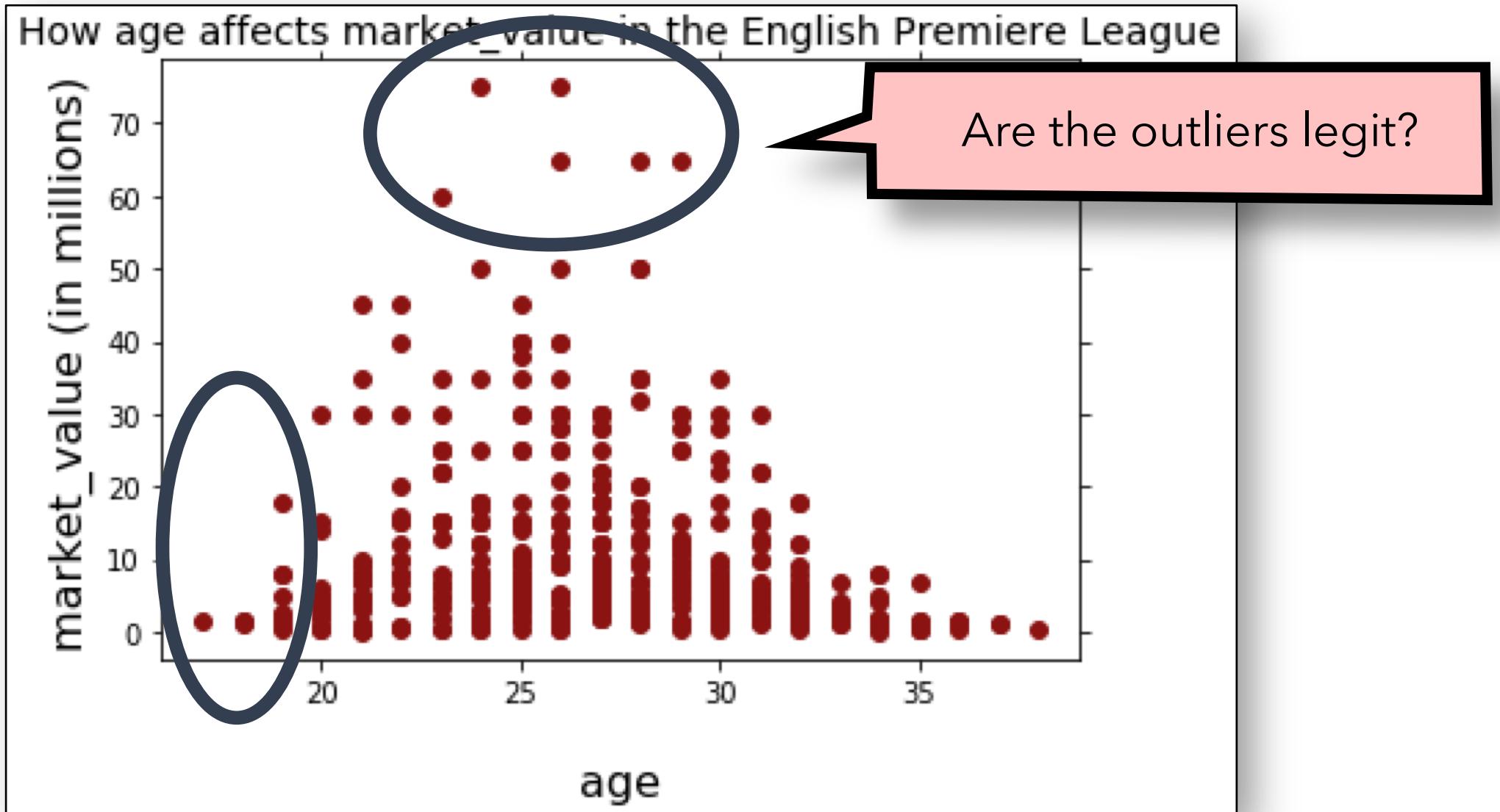
EDA Refresher: English Premier League



EDA Refresher: English Premier League



EDA Refresher: English Premier League



EDA Refresher: English Premier League

```
league_df.loc[league_df['age']<20][['name', 'club', 'age', 'position', \n    'market_value']].sort_values(by="age")
```

	name	club	age	position	market_value
233	Ben Woodburn	Liverpool	17	LW	1.50
231	Trent Alexander-Arnold	Liverpool	18	RB	1.50
350	Josh Tymon	Stoke+City	18	LB	1.00
435	Jonathan Leko	West+Brom	18	RW	1.50
147	Tom Davies	Everton	19	CM	8.00
155	Ademola Lookman	Everton	19	LW	5.00
239	Dominic Solanke	Liverpool	19	CF	2.00
270	Marcus Rashford	Manchester+United	19	CF	18.00
281	Axel Tuanzebe	Manchester+United	19	CB	1.00
282	Timothy Fosu-Mensah	Manchester+United	19	DM	2.50
375	Tammy Abraham	Swansea	19	CF	8.00
436	Sam Field	West+Brom	19	CM	0.25

EDA Refresher: English Premier League

```
league_df.loc[league_df['market_value']>=60][['name', 'club', \n    'age', 'position', 'market_value']].sort_values(by="age")
```

			name	club	age	position	market_value
377			Harry Kane	Tottenham	23	CF	60.0
263			Paul Pogba	Manchester+United	24	CM	75.0
92			Eden Hazard	Chelsea	26	LW	75.0
240			Kevin De Bruyne	Manchester+City	26	AM	65.0
0			Alexis Sanchez	Arsenal	28	LW	65.0
241			Sergio Aguero	Manchester+City	29	CF	65.0

EDA Refresher: English Premier League

```
league_df.loc[league_df['market_value']>=60][['name', 'club', \n    'age', 'position', 'market_value']].sort_values(by="age")
```

	name	club	age	position	market_value
377	Harry Kane	Tottenham	23	CF	60.0
263	Paul Pogba	Manchester+United	24	CM	75.0
92	Eden Hazard	Chelsea	26	LW	75.0
240	Kevin De Bruyne	Manchester+City	26	AM	65.0
0	Alexis Sanchez	Arsenal	28	LW	65.0
241	Sergio Aguero	Manchester+City	29	CF	65.0

Agenda

- EDA Refresher
- Effective Visualization
 - Graphical Integrity
 - Scope
 - Displays
 - Sensible Design
- Communication
 - Motivation
 - Key Considerations

DISCLAIMER.

Some of these examples involve political data. In no way should this be taken as a signal of my support or endorsement in any beliefs; the point is merely to convey good and bad choices when it comes to effective visualization.

Agenda

- EDA Refresher
- Effective Visualization
 - Graphical Integrity
 - Scope
 - Displays
 - Sensible Design
- Communication
 - Motivation
 - Key Considerations

DISCLAIMER.

Some of these examples involve political data. In no way should this be taken as a signal of my support or endorsement in any beliefs; the point is merely to convey good and bad choices when it comes to effective visualization.

Agenda

- EDA Refresher
- Effective Visualization
 - Graphical Integrity
 - Scope
 - Displays
 - Sensible Design
- Communication
 - Motivation
 - Key Considerations

DISCLAIMER.

Some of these examples involve political data. In no way should this be taken as a signal of my support or endorsement in any beliefs; the point is merely to convey good and bad choices when it comes to effective visualization.