Big Data Analytics
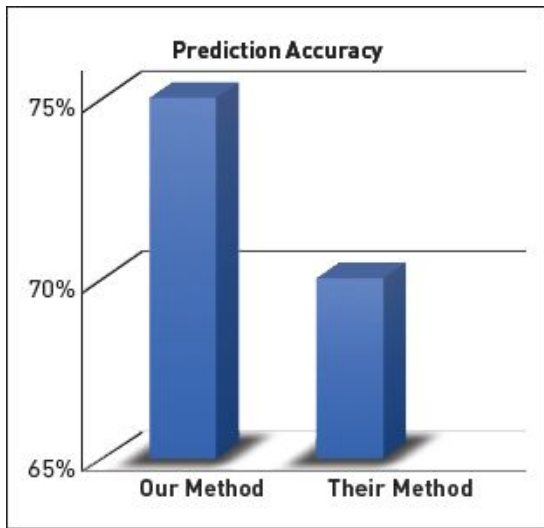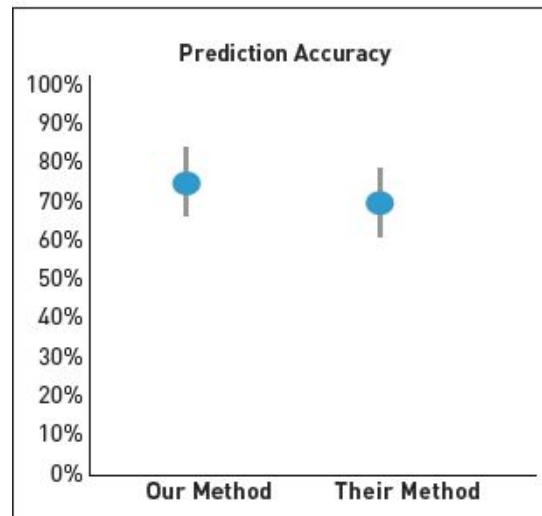
Examples of good and bad data visualization

Misleading visualization: Proportion

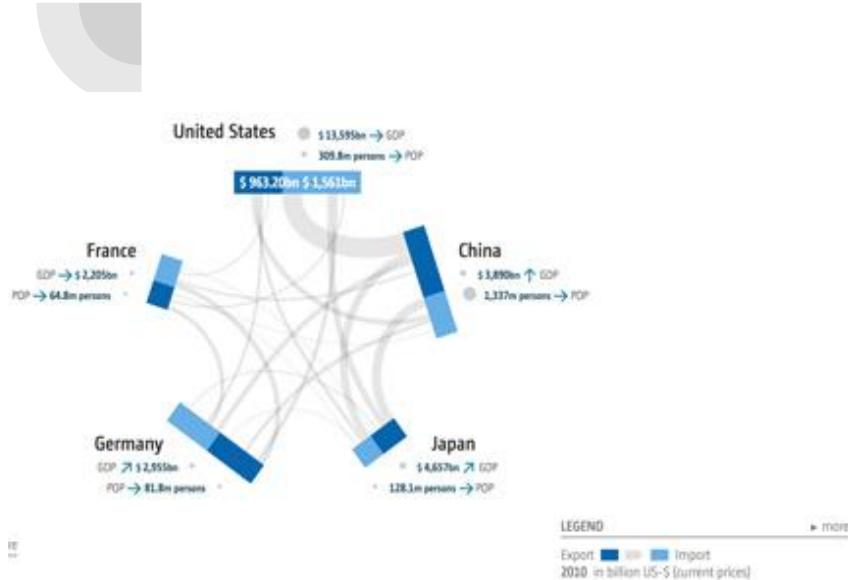
Misleading scale: the method seems twice as accurate

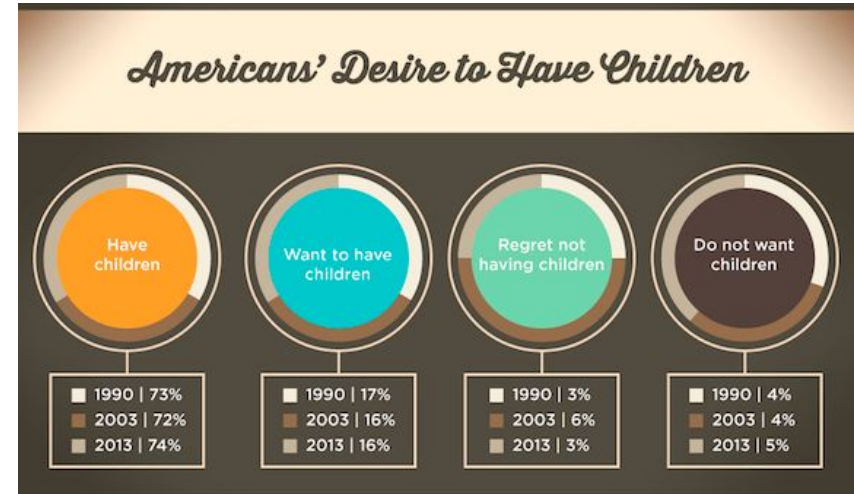**True accuracy improvement: not very significant**

Aesthetic is not enough

United States
○ $13,595bn → GDP
○ 309.8m persons → POP
$ 963.20bn $ 1,561bn

France
GDP → $ 2,205bn
POP → 64.8m persons

China
○ $ 3,890bn ↑ GDP
○ 1,337m persons → POP

Germany
GDP ↗ $ 2,955bn
POP → 81.8m persons

Japan
$ 4,657bn ↗ GDP
128.1m persons → POP

LEGEND   ► more
Export ■ ■ Import
2010 in billion US-$ (current prices)

- **Beautiful**
- **Informative**
- **Consistent scales**
- **Clear message**

Americans' Desire to Have Children

Have children	Want to have children	Regret not having children	Do not want children
■ 1990	73%	■ 1990	17%	■ 1990	3%	■ 1990	4%
■ 2003	72%	■ 2003	16%	■ 2003	6%	■ 2003	4%
■ 2013	74%	■ 2013	16%	■ 2013	3%	■ 2013	5%

- **Numbers don't add up to 100%**
- **inconsistent choice of dates**
- **Poor choice of chart for the given data**
- **Unclear message**

# Football transfers Data set

Data set overview

The dataset of top 250 most expensive football transfers from season 2000-2001 until 2018-2019. It was created on 1 August 2018 and thus may have an incomplete list of the latest transfer window (Summer 2018)

Name <fctr>	Position <fctr>	Age <int>	Team_from <fctr>	League_from <fctr>	Team_to <fctr>	League_to <fctr>	Season <fctr>	Market_value <int>	Transfer_fee <int>
Lu\303\255s Figo	Right Winger	27	FC Barcelona	LaLiga	Real Madrid	LaLiga	2000–2001	NA	60000000
Hern\303\241n Crespo	Centre–Forward	25	Parma	Serie A	Lazio	Serie A	2000–2001	NA	56810000
Marc Overmars	Left Winger	27	Arsenal	Premier League	FC Barcelona	LaLiga	2000–2001	NA	40000000
Gabriel Batistuta	Centre–Forward	31	Fiorentina	Serie A	AS Roma	Serie A	2000–2001	NA	36150000
Nicolas Anelka	Centre–Forward	21	Real Madrid	LaLiga	Paris SG	Ligue 1	2000–2001	NA	34500000
Rio Ferdinand	Centre–Back	22	West Ham	Premier League	Leeds	Premier League	2000–2001	NA	26000000

This Data Set has 18 possible values for the position therefore we decided to regroup them in 3 categories to make the analysis pertinent : Attack, Defense, Midfield
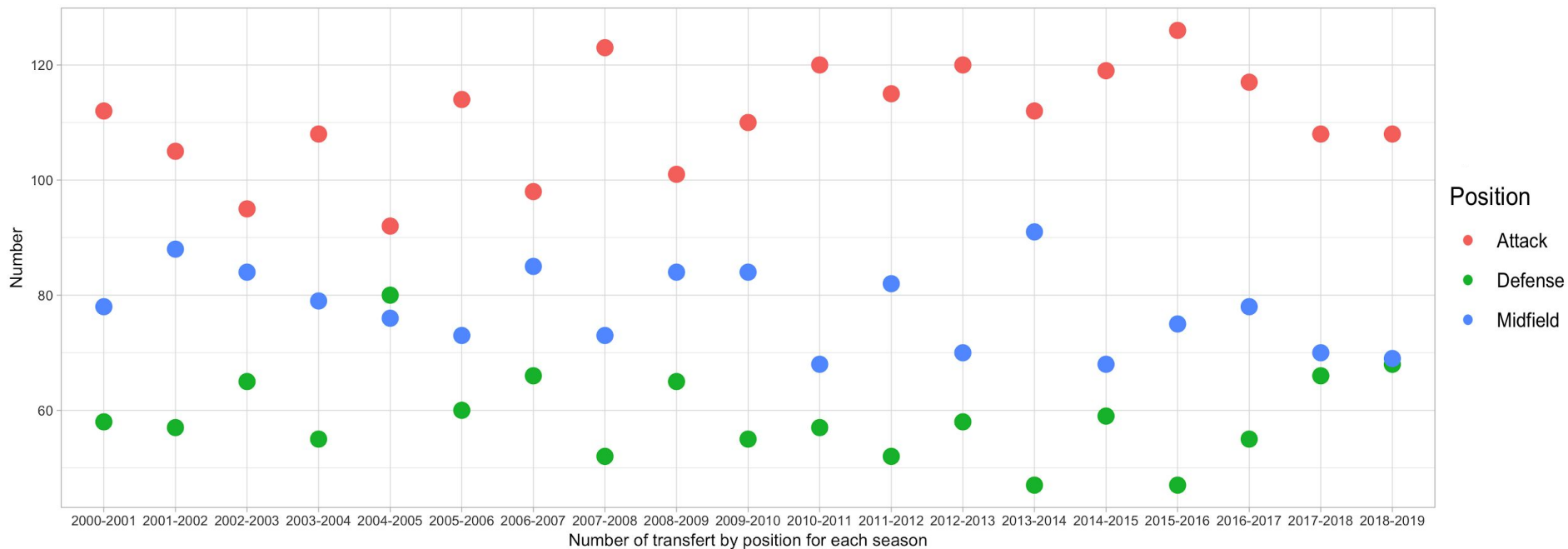
Link: https://www.kaggle.com/vardan95ghazaryan/top-250-football-transfers-from-2000-to-2018
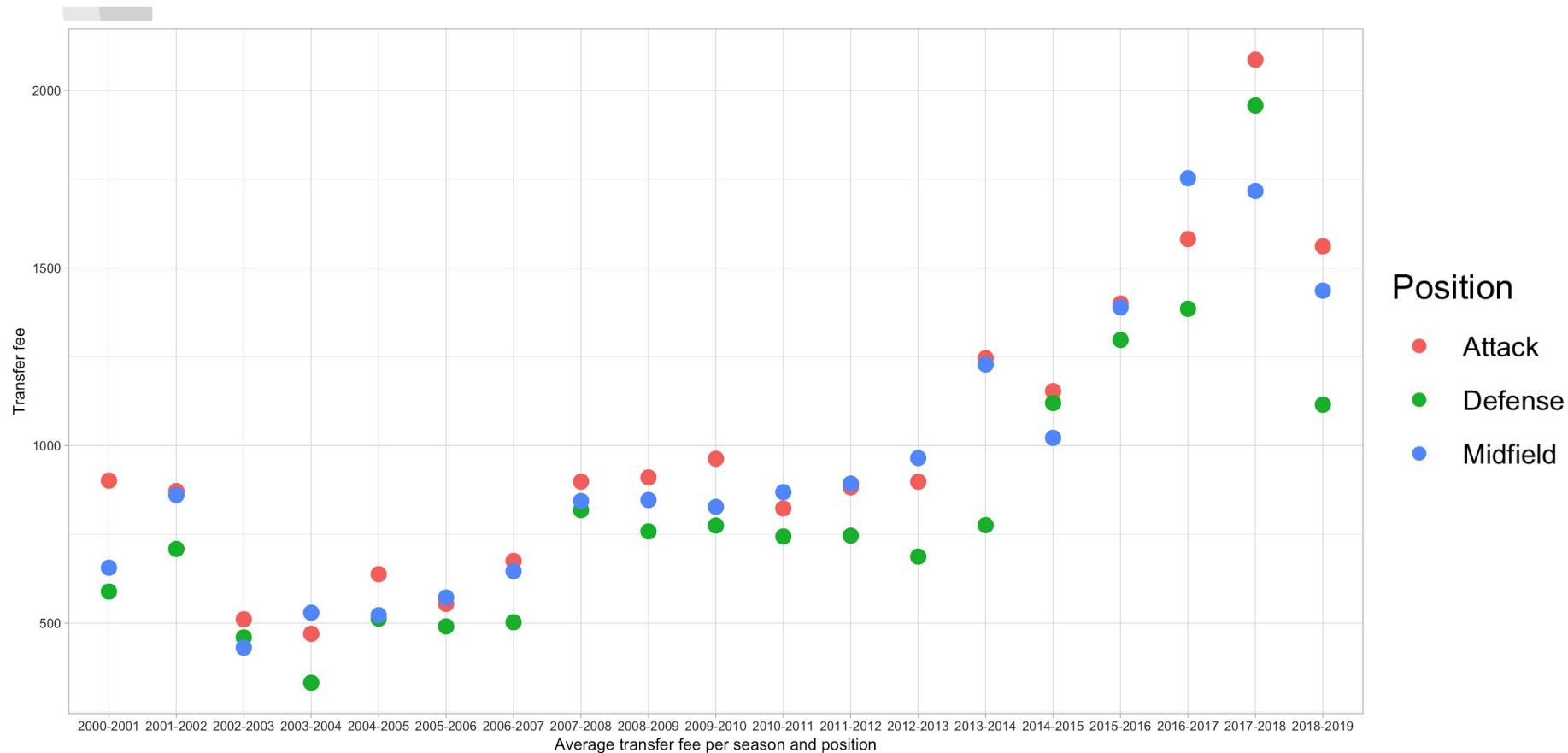
Average age when transferred
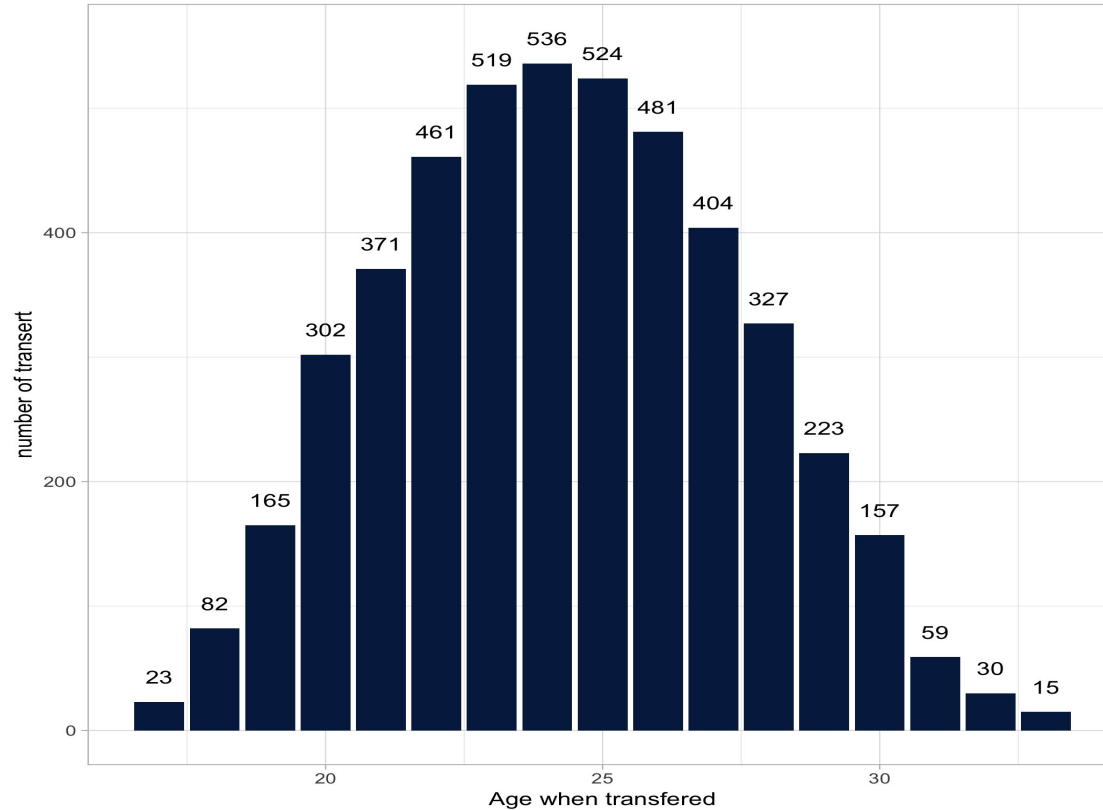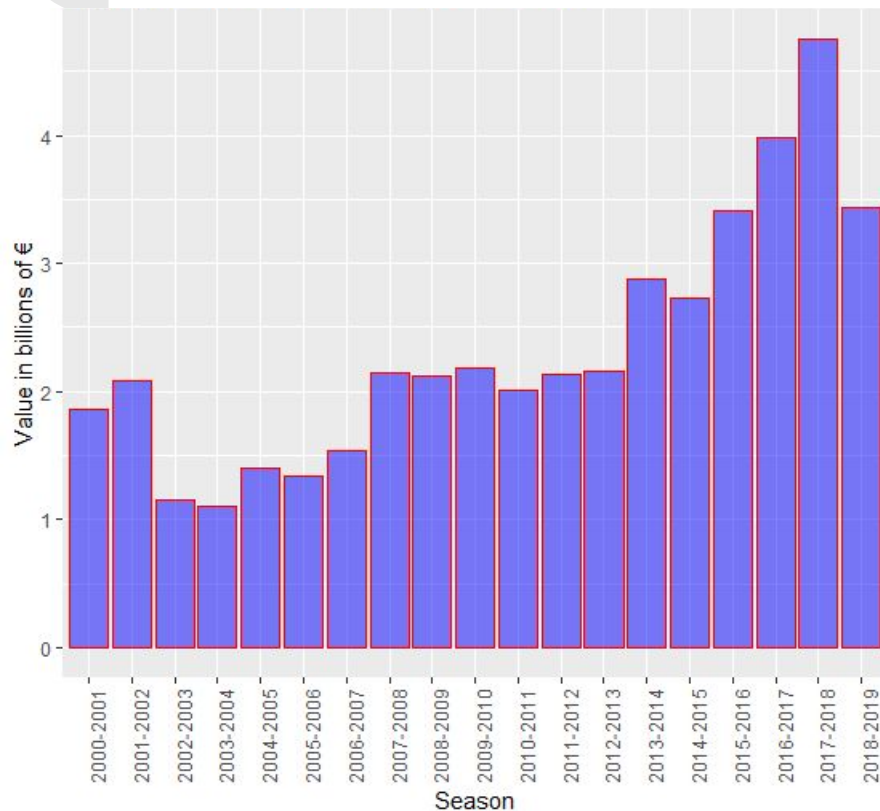
Number of transfers by position for each season
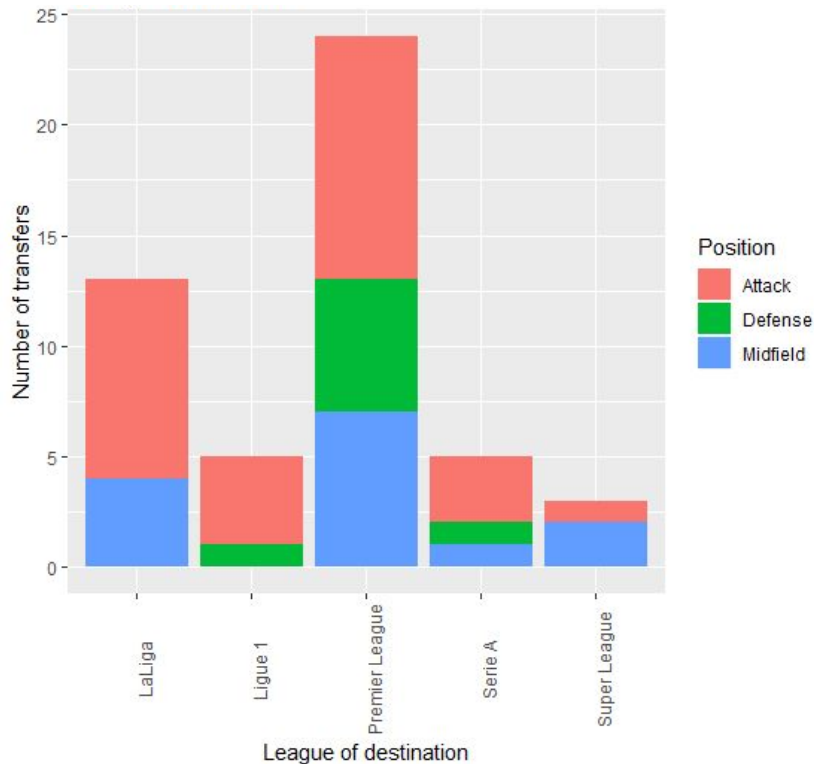
Average transfer fee per season and position

Aggregate value of transfers per season


❖ A Multi-billion $ market that is increasing

❖ Aggregate value in 2018-2019 seems lower as it is incomplete: Dataset ends up on August 1st while the transfer window ended up on August 31st)

The 50 most expensive transfers in all leagues

- ❖ **Concentration of money and talent in a few european leagues**
  - ➢ *Among the 65 Leagues of our dataset, only 5 are represented on the 50 most expensive transfers*

- ❖ **Financial domination of two leagues: The Premier League and La Liga**

- ❖ **More than half of the most expensive transfers involve attacking players**

The 50 most expensive transfers in Ligue 1
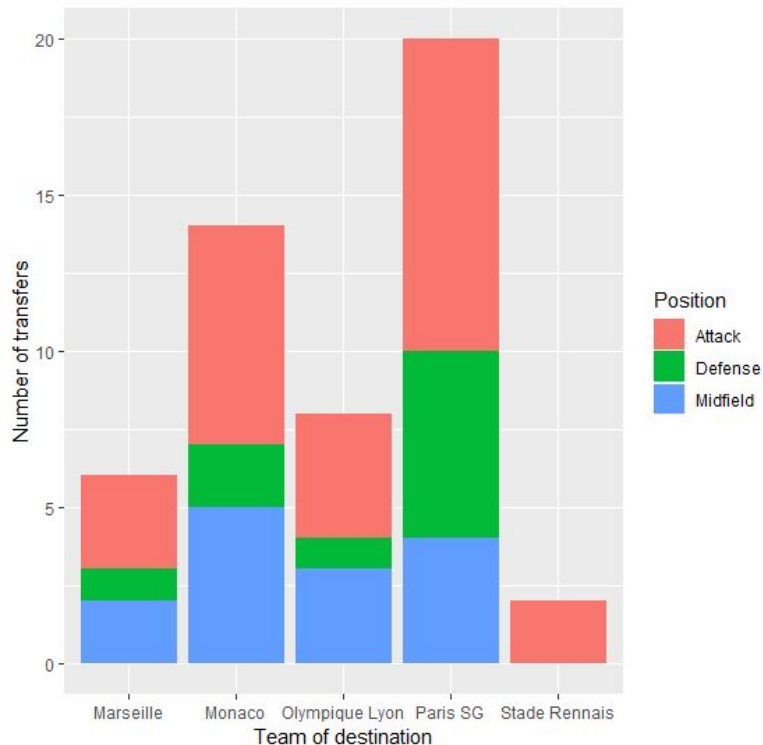
❖ **Financial domination of a few clubs**

➢ Among the 20 in the league, only 5 were involved in the 50 most expensive deals

❖ **Paris Saint-Germain and Monaco are largely ahead in terms of the number of expensive deals**

➢ **Teams with foreign investments**
Russian billionaire Dimitri Rybolovlev for Monaco, Qatar state for PSG)

Fees of the 50 most expensive transfers of Ligue per season (in M of €)
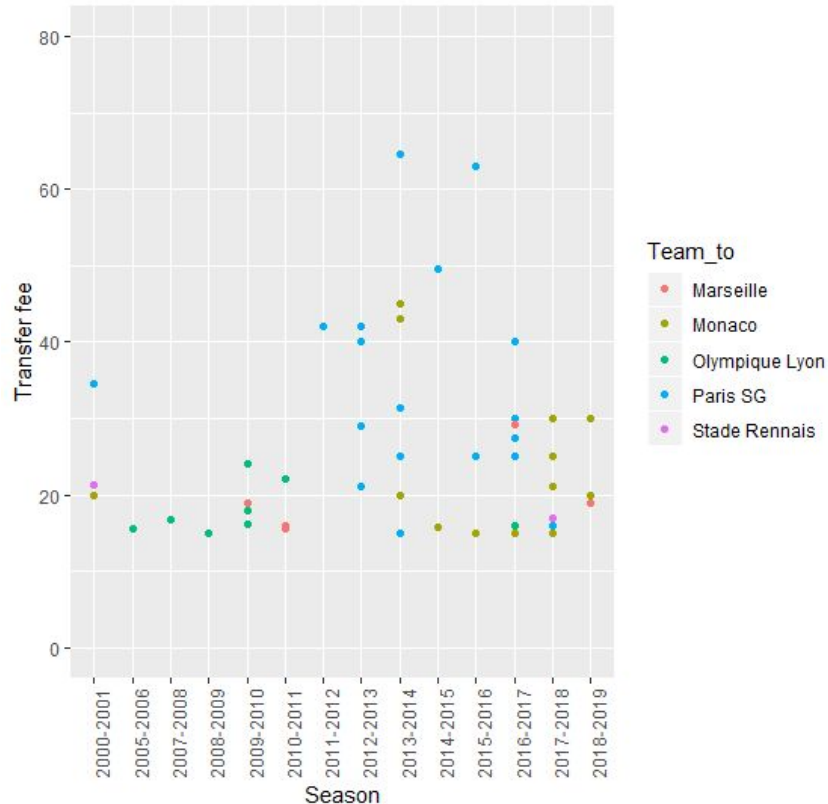
❖ **Lots of Green dots from 2000 to 2010**

➢ Domination of the league by Olympique Lyonnais (7 consecutive titles from 2002 to 2008)

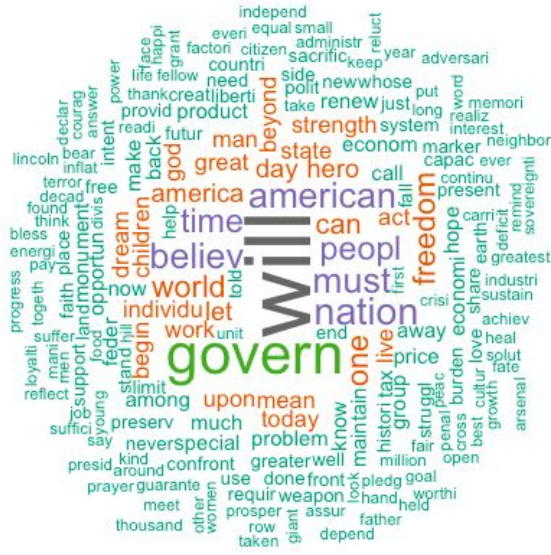❖ **Lots of high Blue dots as of 2011**

➢ New Qatari fund bought PSG in 2011 and domination of the club since then

Word Cloud –
US Presidents Inaugural Speeches

Word Cloud - Inaugural Speeches

Kennedy - 1961   Reagan - 1981   Obama - 2009

Word Cloud - Inaugural Speeches

word	freq
let	16
power	9
can	9
world	8
nation	8
side	8
new	7
pledg	7
will	7
ask	6

Democrat

word	freq
will	33
govern	19
nation	11
must	10
american	10
believ	10
peopl	9
time	9
world	8
one	8

Republican

word	freq
will	19
nation	15
can	13
applaus	12
new	11
america	10
generat	9
everi	8
peopl	8
must	8

Democrat

Group / Analysis	Values	Word Cloud
Republican	Based on **individual** rights and justice. *Conservative*	*Time*, must, **one**
Democrat	Based on **community** and social responsibility. *Progressive*	*New*, can, **every**

General Trends : Let, ask → Must, will
World → Nation, America

Word Cloud - Inaugural Speeches

Trump - 2017

	word	freq
will	will	40
america	america	17
american	american	16
countri	countri	12
nation	nation	10
peopl	peopl	9
one	one	9
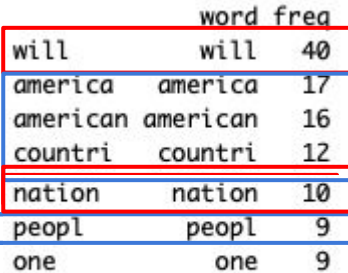
Word Cloud - The Method

Pros : Engaging, Fast, Reveals the essential

Cons : Size versus length of word can be misleading, and context may be important