



The R Environment for Performance Analysis

Wal McConnell
wal@backfour.com
[@walmc](https://twitter.com/walmc)

AGENDA

- Background
- Data Science Process
- Programming
- R
- Integrating Sport Science Data





Background



SPAIN



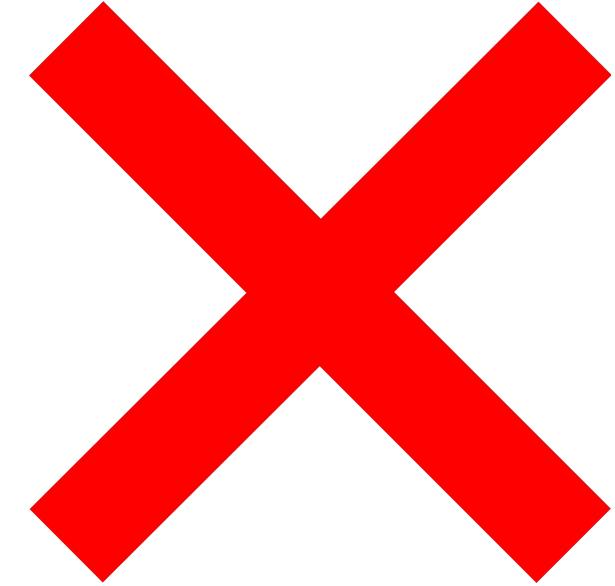






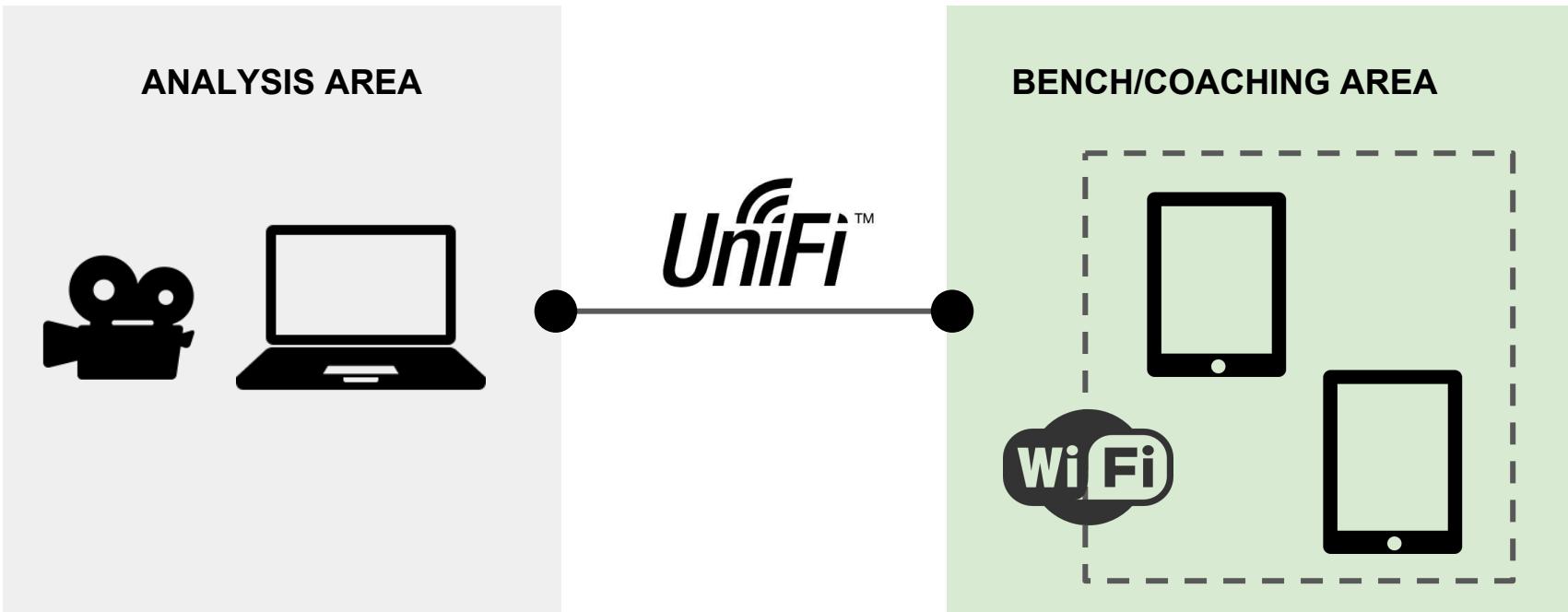
Hockey Ireland PA Process

- **Pre tournament**
 - Opposition Analysis (Set Piece)
 - Objective setting
- **In tournament**
 - Objectives measurement
 - In game stats & video
 - Match reporting (us and opposition)
 - Tournament reporting
- **Post tournament**
 - Reporting, databasing
 - Feedback



SPORTS CODE

Stadium Setup

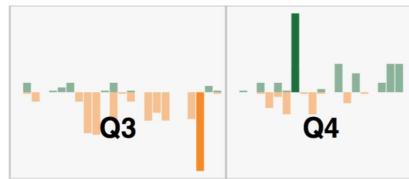


M3 - Ireland v Malaysia

Azlan Shah - Ipoh, Malaysia

3rd March 2018

Team	Goals			Penalty Corners		Shots		
	Goal	Field Goal	PC Goal	PC	PC %	Shots	On Target	Off Target
IRE	1	1	0	5	0%	5	4	1
MAL	4	2	2	7	28.6%	16	14	2



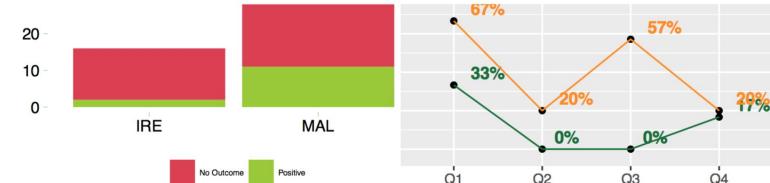
With Ball Objectives

Description	
50% Circle Entry => Outcome	16 / 2 (12.5%)
Long Corner / Handball Outcomes	15.4%
Counter - to Score / Keep ball in final third	
Field Goals	1
PCA %	0%

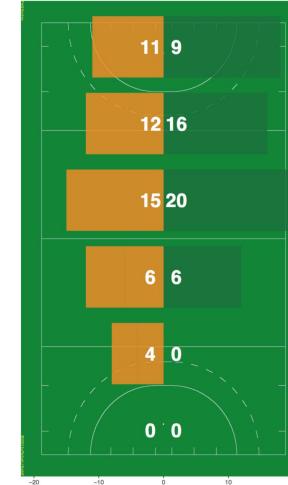
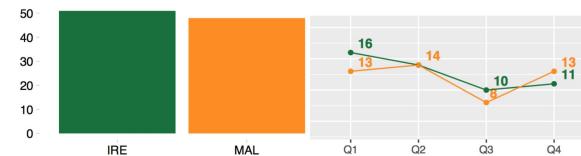
Without Ball Objectives

Description	
Concede no goals in first/last 2 minutes	4th goal (Stroke) , -2 mins before end of 2nd Q
On-Field communication / Route 1	
PCD %	28.6%
Circle Entries / Opposition Circle Entry outcomes	28 / 11 (39.3%)

Circle Entry Outcomes



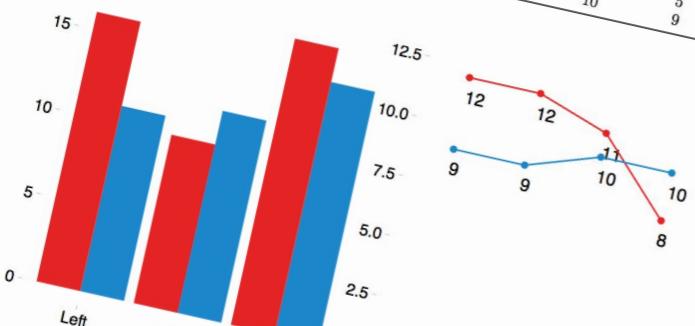
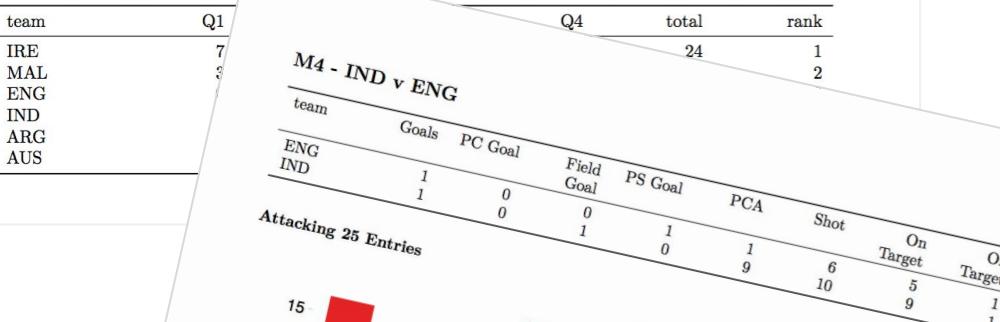
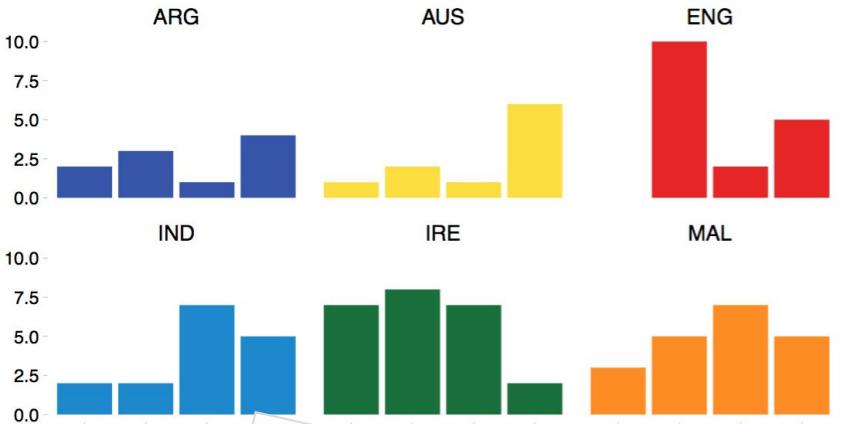
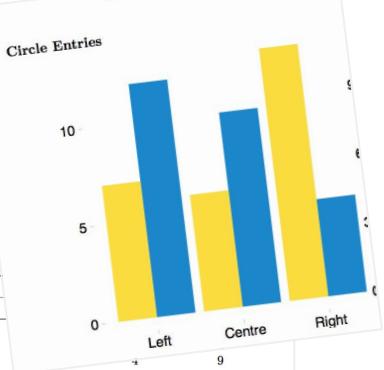
Turnovers



Azlan Shah - Tournament Report
Ipoh, Malaysia
7th March 2018

Pool Matches

team	Goals		
	Goals	Field Goal	PC Goal
M1-ARG-IND	3	0	3
IND	2	0	2
M2-AUS-ENG	4	4	0
AUS	1	1	0
ENG	0	3	12
M3-IRE-MAL	1	1	0
IRE	4	2	2
MAL	0	7	16
M4-IND-ENG	1	0	0
ENG	1	1	0
IND	0	9	10
M5-IRE-ARG	5	1	3
IRE	3	3	0
ARG	1	4	6
M6-AUS-MAL	3	2	1
AUS	1	1	0
MAL	0	1	2
M7-IRE-ENG	4	3	1
IRE	1	0	1
ENG	0	7	13
M8-AUS-IND	4	2	1
AUS	2	2	0
IND	0	2	5
M9-ARG-MAL	1	1	0
MAL	2	2	0
ARG	0	6	11
M10-IRE-AUS	4	3	1
IRE	1	0	1
AUS	0	0	4
M11-ARG-ENG	1	0	0
ENG	1	0	1
ARG	0	4	5
M12-IND-MAL	5	3	2
IND	1	0	1
MAL	0	5	12





KITMAN LABS



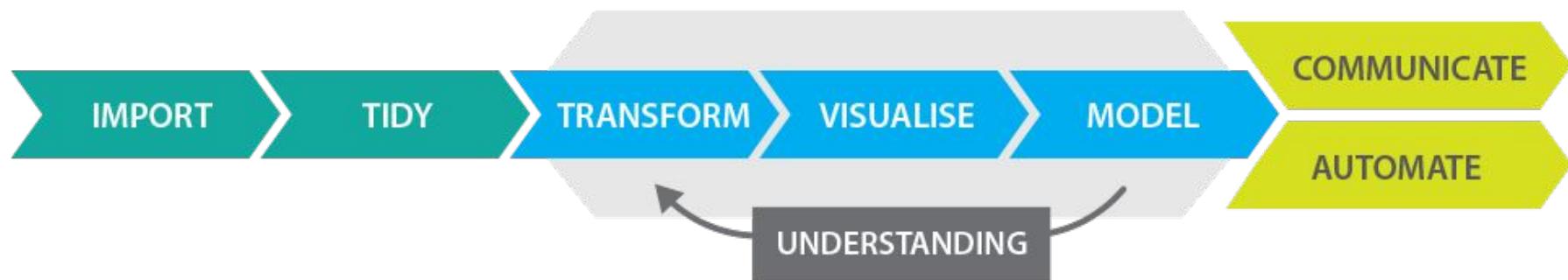
KITMAN LABS





Data Science Process

Data Science Process



“R4DS Data Science Process”
Hadley Wickham

IMPORT

TIDY

TRANSFORM

VISUALISE

MODEL

COMMUNICATE

AUTOMATE



1. Each variable you measure should be in one column.
2. Each different observation of that variable should be in a different row.
3. There should be one table for each "kind" of variable.
4. If you have multiple tables, they should include a column in the table that allows them to be linked.

IMPORT

TIDY

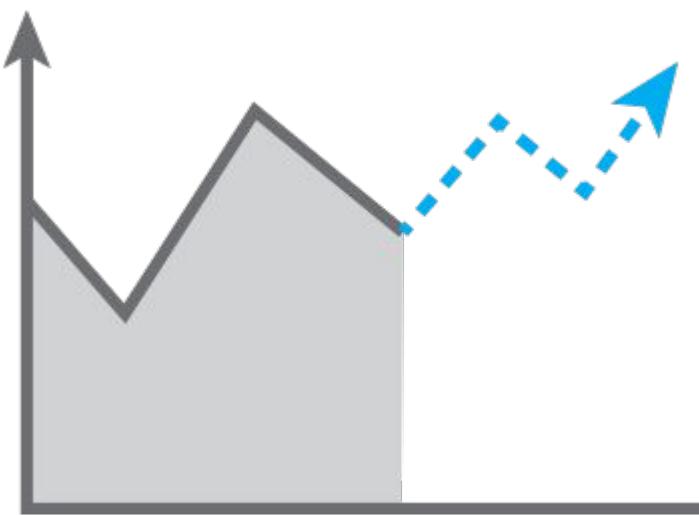
TRANSFORM

VISUALISE

MODEL

COMMUNICATE

AUTOMATE



IMPORT

TIDY

TRANSFORM

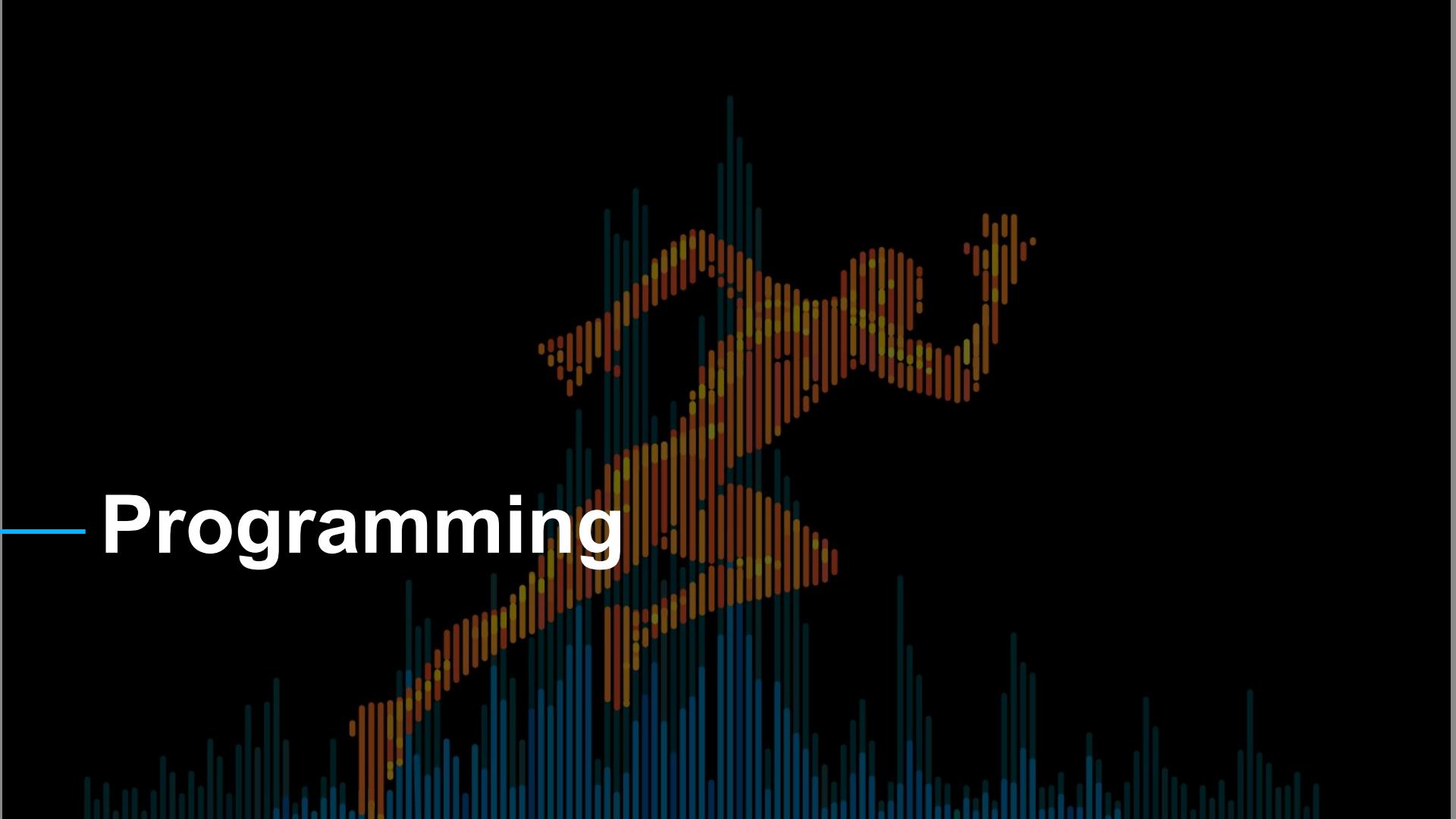
VISUALISE

MODEL

COMMUNICATE

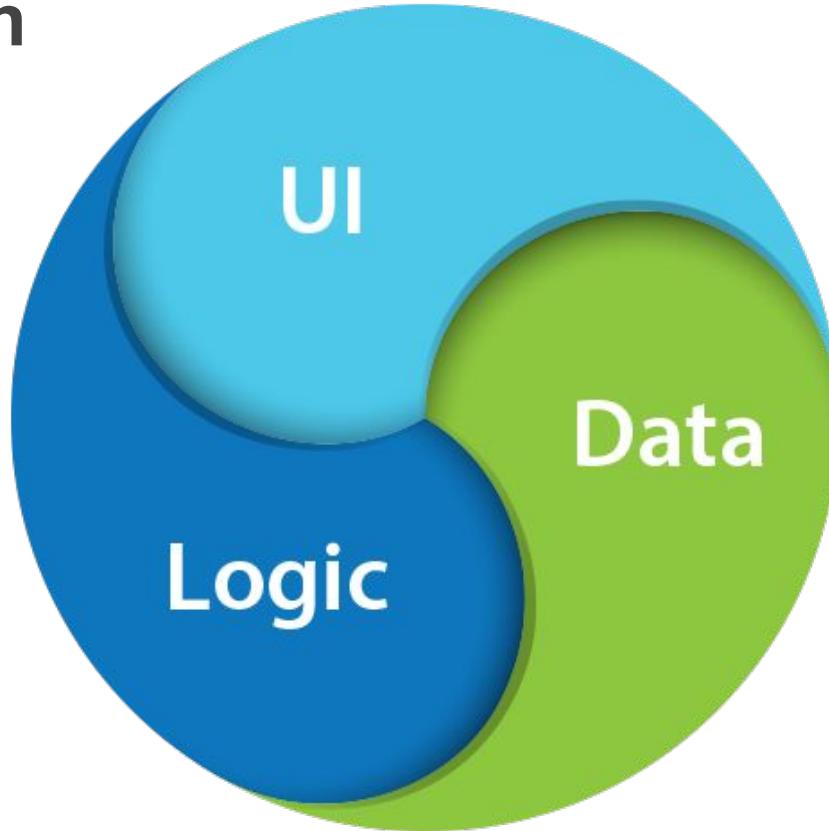
AUTOMATE





— Programming

Abstraction



```
dens <- density(data, n = npts)
dx <- dens$x
dy <- dens$y
if(add == TRUE)
  plot(0., 0,
       ylab = "Density")
if(orientation == "vertical")
  dx2 <- (dx - min(dx)) / max(dx)
  x[1.]
  dy2 <- (dx - min(dy)) / max(dy)
  y[1.]
  seqbelow <- rep(y[1.], length(dx))
  if(Fill == T)
    confshade(dx2, seqbelow, dy2)
```

Text



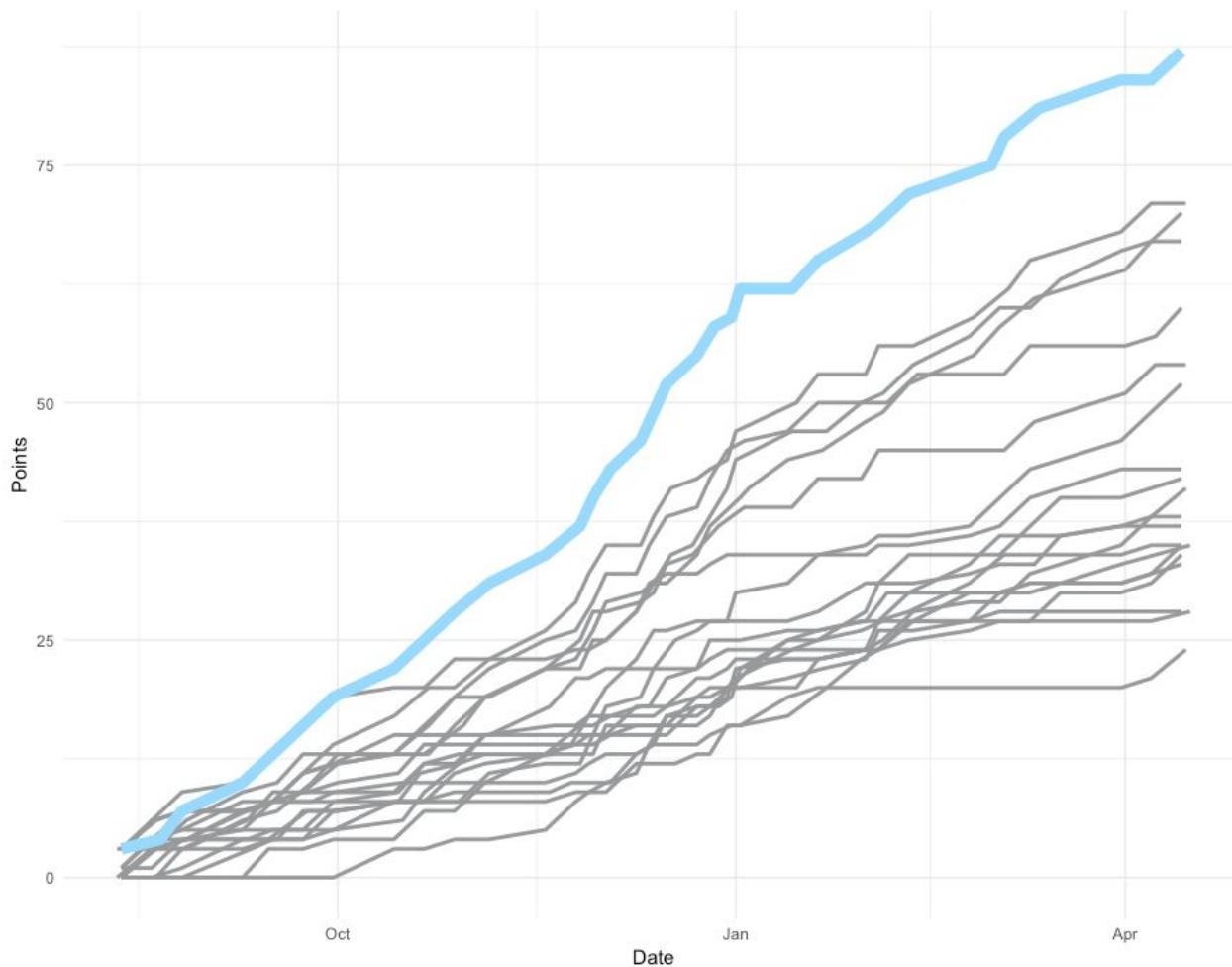
Version Control



R



Man City - 2017/2018 points accumulation



```

# Step 1 : Import Data
premier_league_2017_2018_url <- "http://www.football-data.co.uk/mmz4281/1718/E0.csv"
data <- read.csv(premier_league_2017_2018_url)

# filter columns to just those we are interested in
data <- data %>% select(HomeTeam, AwayTeam, Date, FTR)

# Convert Date column (DD/MM/YYYY)
data$date <- as.Date(data$date, format="%d/%m/%y")

```

	HomeTeam	AwayTeam	Date	FTR
1	Arsenal	Leicester	2017-08-11	H
2	Brighton	Man City	2017-08-12	A
3	Chelsea	Burnley	2017-08-12	A
4	Crystal Palace	Huddersfield	2017-08-12	A
5	Everton	Stoke	2017-08-12	H
6	Southampton	Swansea	2017-08-12	D
7	Watford	Liverpool	2017-08-12	D
8	West Brom	Bournemouth	2017-08-12	H
9	Man United	West Ham	2017-08-13	H
10	Newcastle	Tottenham	2017-08-13	A
11	Bournemouth	Watford	2017-08-19	A

```

# Step 2 : Tidy Data
home_data <- data %>%
  select(HomeTeam, Date, FTR) %>%
  mutate(Points = case_when(
    .FTR == 'H' ~ 3,
    .FTR == 'D' ~ 1,
    .FTR == 'A' ~ 0)) %>%
  rename(Team = HomeTeam)

away_data <- data %>%
  select(AwayTeam, Date, FTR) %>%
  mutate(Points = case_when(
    .FTR == 'H' ~ 0,
    .FTR == 'D' ~ 1,
    .FTR == 'A' ~ 3)) %>%
  rename(Team = AwayTeam)

# Combine home and away data
season_data <- rbind(home_data, away_data) %>% arrange(Date)

```

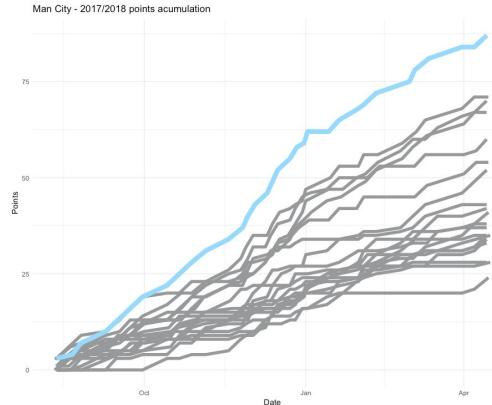
Team	Date	FTR	Points
Arsenal	2017-08-11	H	3
Leicester	2017-08-11	H	0
Brighton	2017-08-12	A	0
Chelsea	2017-08-12	A	0
Crystal Palace	2017-08-12	A	0
Everton	2017-08-12	H	3
Southampton	2017-08-12	D	1
Watford	2017-08-12	D	1
West Brom	2017-08-12	H	3
Man City	2017-08-12	A	3
Burnley	2017-08-12	A	3
Huddersfield	2017-08-12	A	3
Stoke	2017-08-12	H	0
Swansea	2017-08-12	D	1
Liverpool	2017-08-12	D	1
Bournemouth	2017-08-12	H	0
Man United	2017-08-13	H	3

```
# Step 3 - Transform Data
season_data_cumulative <- season_data %>%
  group_by(Team) %>%
  mutate(cumulative_points = cumsum(Points))
```

Team	Date	FTR	Points	cumulative_points
Arsenal	2017-08-11	H	3	3
Leicester	2017-08-11	H	0	0
Brighton	2017-08-11		0	0
Chelsea	2017-08-12	A	0	0
Crystal Palace	2017-08-12	A	0	0
Everton	2017-08-12	H	3	3
Southampton	2017-08-12	D	1	1
Watford	2017-08-12	D	1	1
West Brom	2017-08-12	H	3	3
Man City	2017-08-12	A	3	3
Burnley	2017-08-12	A	3	3
Huddersfield	2017-08-12	A	3	3
Stoke	2017-08-12	H	0	0
Swansea	2017-08-12	D	1	1
Liverpool	2017-08-12	D	1	1
Bournemouth	2017-08-12	H	0	0
Man United	2017-08-13	H	3	3
Newcastle	2017-08-13	A	0	0
West Ham	2017-08-13	H	0	0
Tottenham	2017-08-13	A	3	3
Bournemouth	2017-08-19	A	0	0
Burnley	2017-08-19	A	0	3
Leicester	2017-08-19	H	3	3
Liverpool	2017-08-19	H	3	4
Southampton	2017-08-19	H	3	4
Stoke	2017-08-19	H	3	3

```
# Step 4 - Visualise Data
man_city_data <- season_data_cumulative %>% filter(Team == 'Man City')

ggplot(season_data_cumulative, aes(Date, cumulative_points, group = Team)) +
  geom_line(color = "#98999B", size = 2) +
  geom_line(data = man_city_data, color = "#95D9FE", size = 3) +
  theme_minimal() +
  labs(y = "Points", x = "Date") +
  ggtitle("Man City - 2017/2018 points accumulation")
```



```

# Step 1 : Import Data
premier_league_2017_2018_url <- "http://www.football-data.co.uk/mmz4281/1718/E0.csv"
data <- read.csv(premier_league_2017_2018_url)

# filter columns to just those we are interested in
data <- data %>% select(HomeTeam, AwayTeam, Date, FTR)

# Convert Date column (DD/MM/YYYY)
data$Date <- as.Date(data$Date, format="%d/%m/%y")

# Step 2 : Tidy Data
home_data <- data %>%
  select(HomeTeam, Date, FTR) %>%
  mutate(Points = case_when(
    .FTR == 'H' ~ 3,
    .FTR == 'D' ~ 1,
    .FTR == 'A' ~ 0)) %>%
  rename(Team = HomeTeam)

away_data <- data %>%
  select(AwayTeam, Date, FTR) %>%
  mutate(Points = case_when(
    .FTR == 'H' ~ 0,
    .FTR == 'D' ~ 1,
    .FTR == 'A' ~ 3)) %>%
  rename(Team = AwayTeam)

# Combine home and away data
season_data <- rbind(home_data, away_data) %>% arrange(Date)

# Step 3 - Transform Data
season_data_cumulative <- season_data %>%
  group_by(Team) %>%
  mutate(cumulative_points = cumsum(Points))

# Step 4 - Visualise Data
man_city_data <- season_data_cumulative %>% filter(Team == 'Man City')

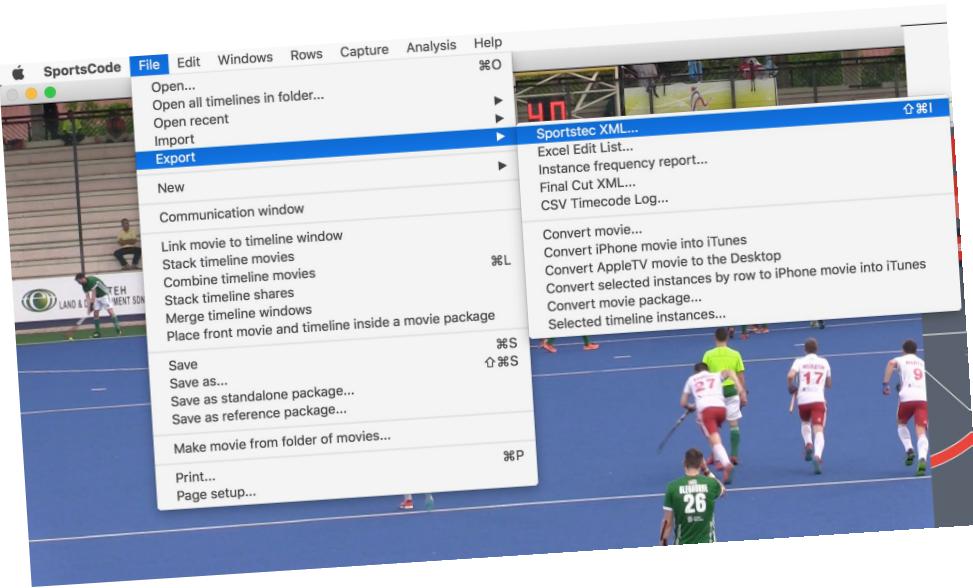
ggplot(season_data_cumulative, aes(Date, cumulative_points, group = Team)) +
  geom_line(color = "#F0F0F0", size = 2) +
  geom_line(data = man_city_data, color = "#95D9FE", size = 3) +
  theme_minimal() +
  labs(y = "Points", x = "Date") +
  ggtitle("Man City - 2017/2018 points accumulation")

```

Analysis Pipeline Approach



Sportscode XML



```
<file>
<SESSION_INFO>
<start_time>2018-03-09 18:05:25.50 +0800</start_time>
</SESSION_INFO>
<ALL_INSTANCES>
<instance>
<ID>4</ID>
<start>134.3963293770</start>
<end>136.3963293770</end>
<code>Match Start/Stop</code>
</instance>
<instance>
<ID>6</ID>
<start>137.5576588010</start>
<end>192.9178699470</end>
<code>IRE Poss.</code>
<label>
<group>Quarter</group>
<text>Q1</text>
</label>
<label>
<group>Attack Entry</group>
<text>Right</text>
</label>
<label>
<group>Outcome</group>
<text>Positive</text>
</label>
</instance>
<instance>
<ID>7</ID>
<start>172.9170916220</start>
<end>183.1570440990</end>
<code>IRE Handball</code>
<label>
<group>Quarter</group>
<text>Q1</text>
</label>
</instance>
<instance>
<ID>8</ID>
<start>175.1570440990</start>
<end>192.3549024800</end>
<code>IRE A25</code>
<label>
<group>Quarter</group>
<text>Q1</text>
</label>
<label>
<group>Attack Entry</group>
<text>Right</text>
</label>
</instance>
<instance>
<ID>9</ID>
<start>177.9178699470</start>
<end>199.1577194590</end>
<code>IRE CP</code>
<label>
<group>CP Number</group>
<text>1st CP</text>
</label>
<label>
<group>Quarter</group>
<text>Q1</text>
</label>
<label>
<group>Circle Pen</group>
<text>Left</text>
</label>
</instance>
```

Possession % stats

```
library(sportscoder)

sportscode_data <- read_sportscode_xml("/Users/wal/Desktop/XML-CARLOW-PRES.xml", format = "tidy")

# Posession %

posession_stats <- sportscode_data %>%
  filter(code %in% c("IRE Poss", "IND Poss")) %>%
  group_by(code) %>%
  summarise(posession_seconds = sum(end - start)) %>%
  mutate(
    posession_percent = posession_seconds / sum(posession_seconds) * 100)
```

```
> posession_stats
# A tibble: 2 x 3
  code      posession_seconds posession_percent
  <fct>          <dbl>            <dbl>
1 IND Poss       1446             44.1
2 IRE Poss       1833             55.9
>
```

Momentum Bar

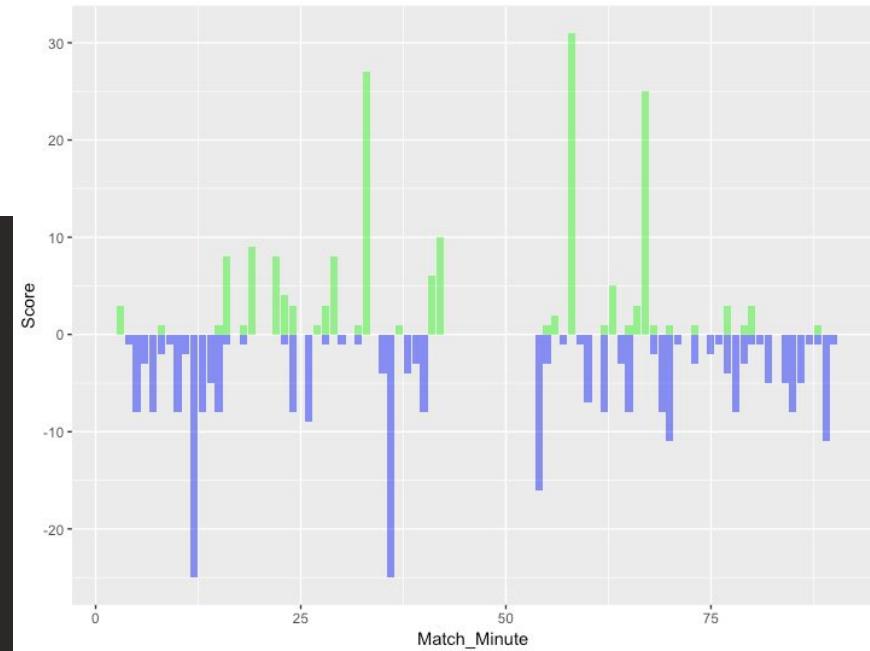
```
# Momentum bar

home_momentum <- match_data %>%
  filter(grepl('IRE', code)) %>%
  group_by(Match_Minute) %>%
  summarise(Score =
    sum(code == "IRE A25") +
    sum(code == "IRE CP") * 2 +
    sum(code == "IRE GSO") * 5 +
    sum(code == "IRE Goal") * 20)

away_momentum <- match_data %>%
  filter(grepl('IND', code)) %>%
  group_by(Match_Minute) %>%
  summarise(Score =
    sum(code == "IND A25") +
    sum(code == "IND CP") * 2 +
    sum(code == "IND GSO") * 5 +
    sum(code == "IND Goal") * 20)

|
home_color = "green"
away_color = "blue"

ggplot() +
  geom_bar(data = home_momentum, aes(Match_Minute, Score), stat = "identity", fill = home_color, alpha = 0.5) +
  geom_bar(data = away_momentum, aes(Match_Minute, -Score), stat = "identity", fill = away_color, alpha = 0.5)
```





R Markdown

R-Markdown

Authoring framework to generate documents from R code



R Markdown

from R Studio

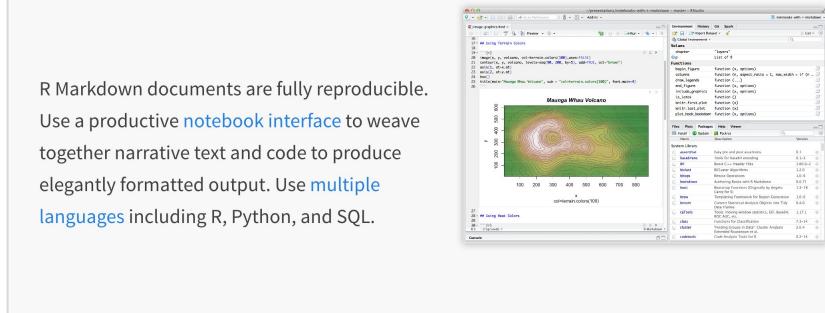
Get Started Gallery Formats Articles



Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown. Turn your analyses into high quality documents, reports, presentations and dashboards.

R Markdown documents are fully reproducible. Use a productive [notebook interface](#) to weave together narrative text and code to produce elegantly formatted output. Use [multiple languages](#) including R, Python, and SQL.



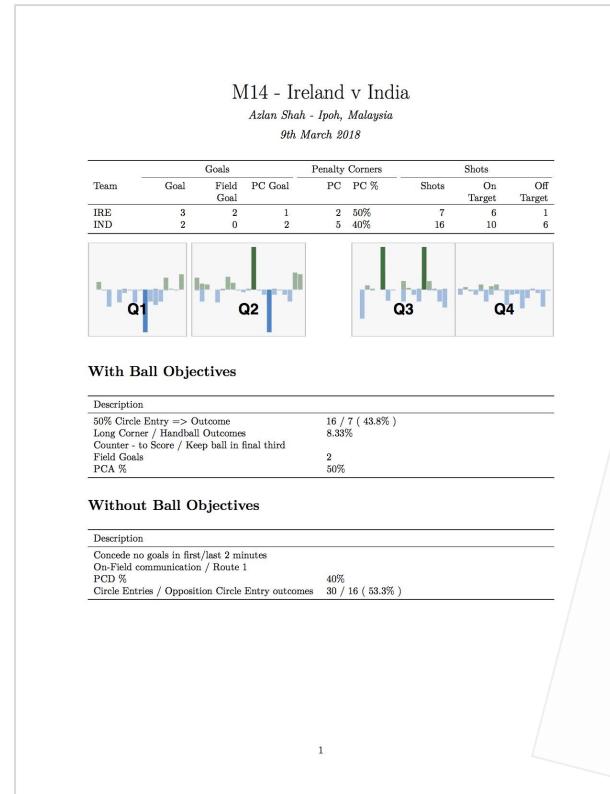
Hockey Ireland - Match Report

RStudio File Edit Code View Plots Session Build D

MatchReport.Rmd *

Knit to HTML Knit to PDF Knit to Word Knit with Parameters... Clear Knitr Cache...

```
1 ti Knit to HTML
2 au Knit to PDF "ayasia"
3 da Knit to Word
4 ou Knit with Parameters...
5
6
7
8
9
10 home_team = 'IRE'
11 away_team = 'IND'
12
13 xml_file_path = "/Users/wal/Desktop/XML-CARLOW-PRES.xml"
14 ...
15
16 ``{r setup, include=FALSE, echo=FALSE}
17 library(data.table)
18 library(knitr)
19 library(kableExtra)
20 library(tidyverse)
21 library(ggthemes)
22 library(scales)
23 library(ggpubr)
24
25 knitr::opts_chunk$set(echo = TRUE)
26 options(knitr.kable.NA = 0)
27 options(knitr.table.format = "latex")
28
29 plot_colors <- c(
```



```

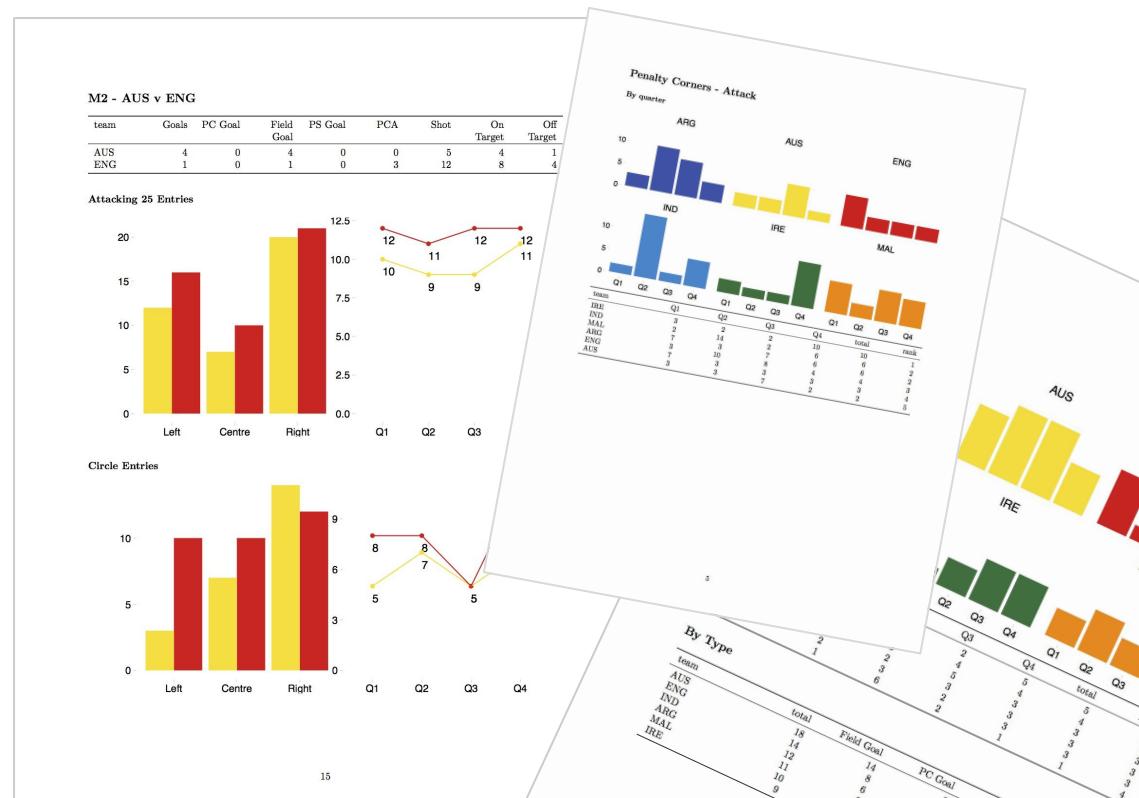
7
8 \newpage
9
0 # Goal Scoring Opportunities
1
2 ````{r,echo=FALSE, out.width='49\\linewidth', fig.width=3.5, fig.height=1.75,fig.show='hold',fig.align='center'}
3 summary <- match_data %>%
4   filter(grepl('GSO', code)) %>%
5   group_by(Team) %>%
6   summarise(Count = n()) %>%
7   filter(!is.na(Team))
8
9 ggplot(summary) +
  geom_bar(aes(Team, Count, fill=Team), stat = "identity", position = "dodge") +
  labs(x = NULL, y = NULL) +
  theme_pander() +
  scale_fill_manual("Team",values = plot_colors) +
  theme(axis.text=element_text(size=10), axis.title=element_text(size=10), legend.position="none")
10
11 quarters_plot(match_data, 'GSO')
12 ````
```



Tournament Report

▼ XML

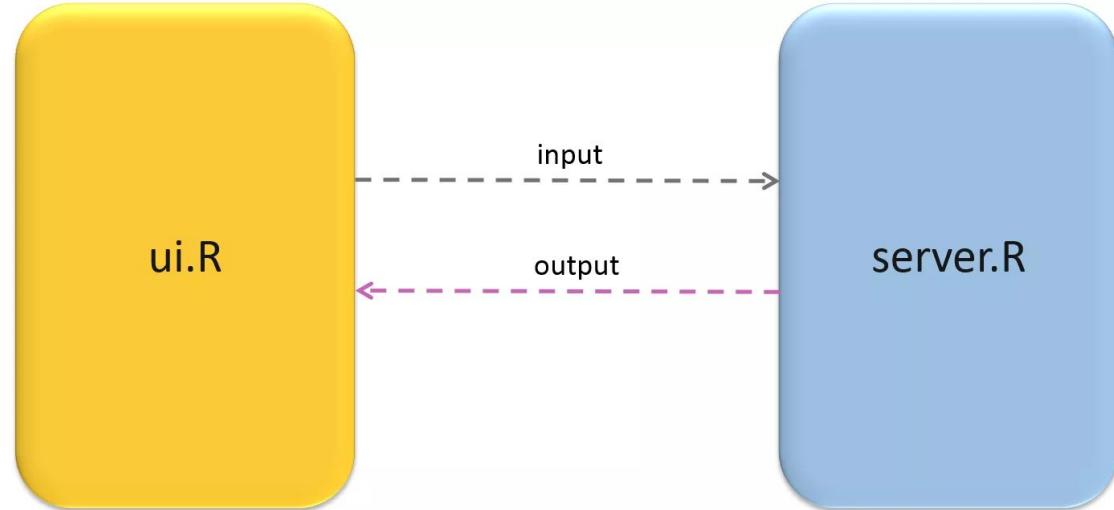
- M1-ARG-IND.xml
- M2-AUS-ENG.xml
- M3-IRE-MAL.xml
- M4-IND-ENG.xml
- M5-IRE-ARG.xml
- M6-AUS-MAL.xml
- M7-IRE-ENG.xml
- M8-AUS-IND.xml
- M9-ARG-MAL.xml
- M10-IRE-AUS.xml
- M11-ARG-ENG.xml
- M12-IND-MAL.xml
- M13-AUS-ARG.xml
- M14-IRE-IND.xml
- M15-MAL-ENG.xml





R Shiny

Shiny is an R system
that makes it easy to
build interactive web
applications straight
from R.





EPL 2017/2018 - Score Prediction

Home Team:

West Brom

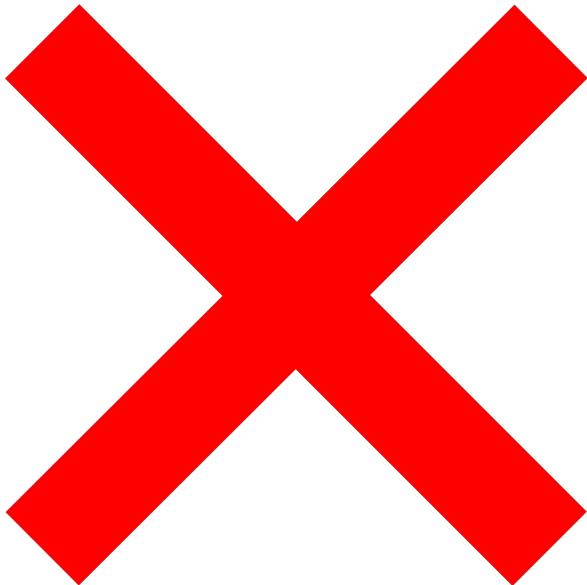
Away Team:

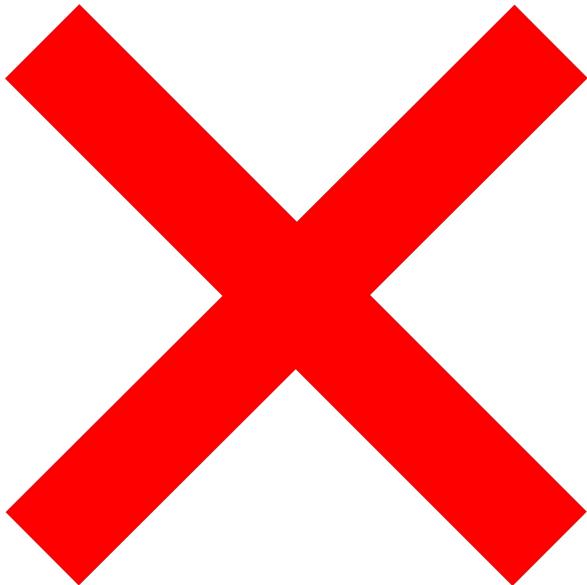
Burnley



Home Win	Draw	Away Win
16.04 %	26.13 %	57.38 %





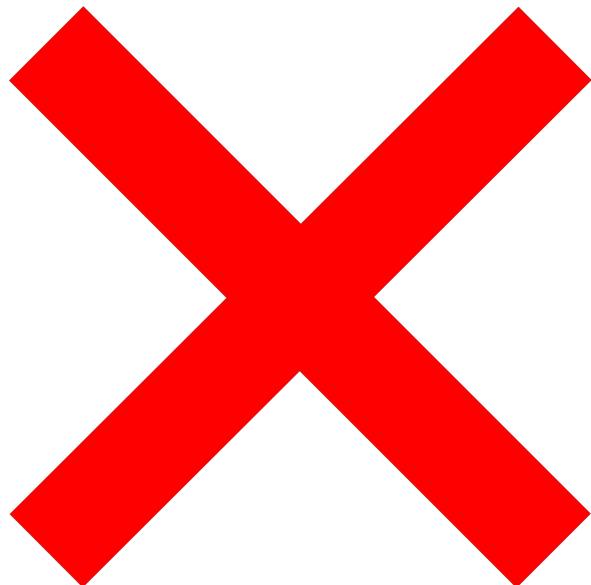


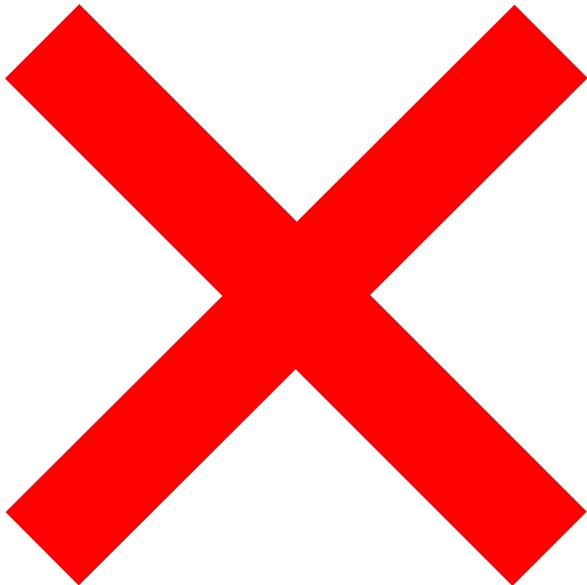


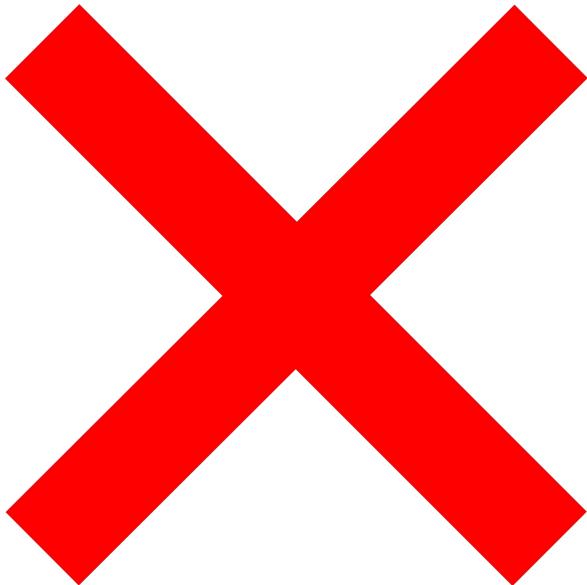
Integrating Sport Science Data

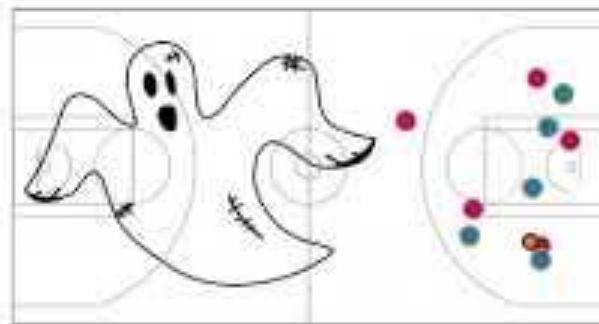
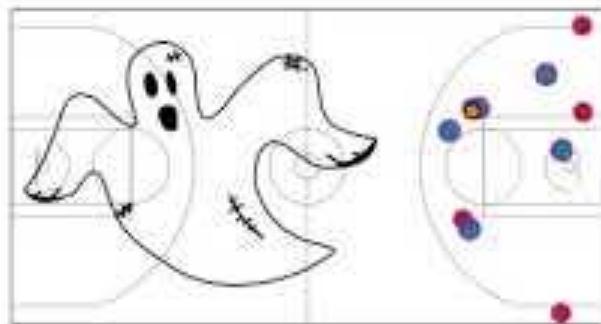
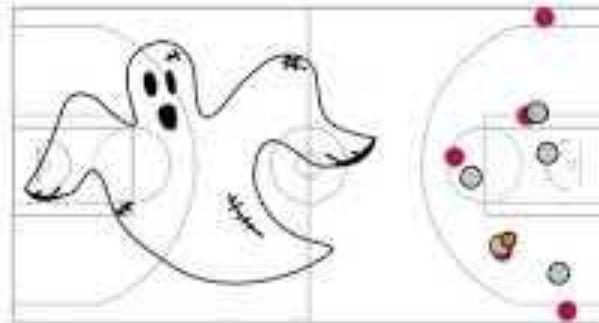
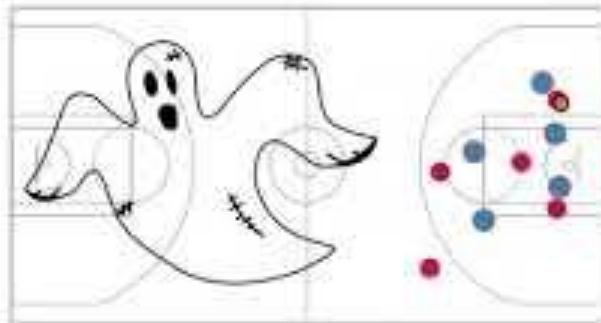
Why?

Meters / Min - v Possession









Books / Resources

R Programming

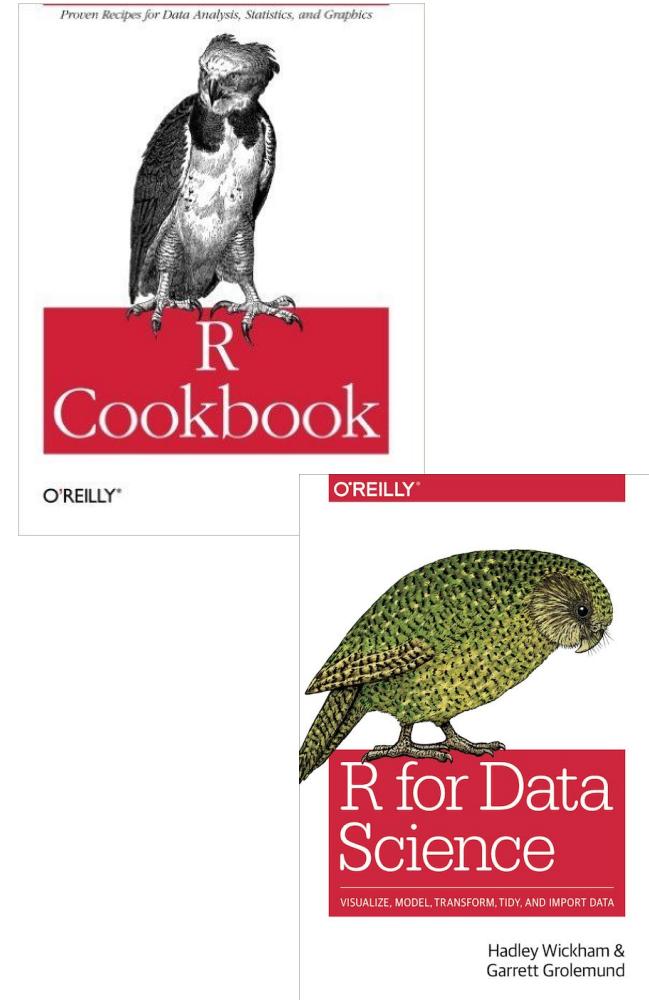
- Tidyverse <https://www.tidyverse.org/>
- R-Markdown <https://rmarkdown.rstudio.com>
- R-Shiny <https://shiny.rstudio.com/>
- R Bloggers <https://www.r-bloggers.com>

[Github.com](#)

[Datacamp.com](#)

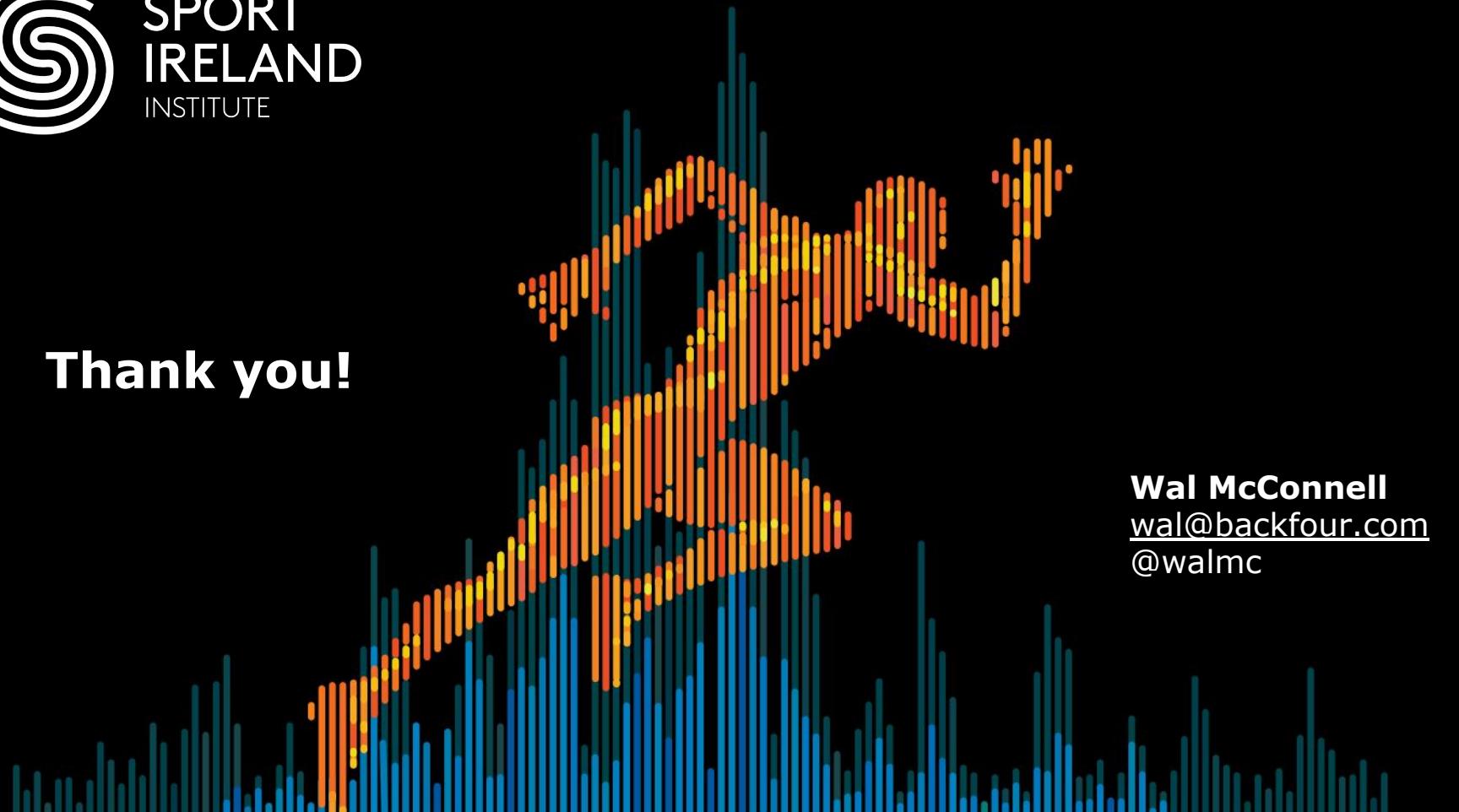
[Stackoverflow.com](#)

Proven Recipes for Data Analysis, Statistics, and Graphics





Thank you!



Wal McConnell
wal@backfour.com
[@walmc](https://twitter.com/walmc)