

# Event Data

Luke Julian

2023-08-05

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
events.wc.22 <- allclean(get.competitionEvents(43,106))
events.wc.18 <- allclean(get.competitionEvents(43,3))
events.wwc.19 <- allclean(get.competitionEvents(72,30))
events.weuro.22 <- allclean(get.competitionEvents(53,106))
```

```
# events.wc.18 <- read.csv("Data/Events/wc22.csv")
# events.wc.22 <- read.csv("Data/Events/wc18.csv")
# events.wwc.19 <- read.csv("Data/Events/wwc19.csv")
# events.weuro.22 <- read.csv("Data/Events/weuro22.csv")
```

```
wc.18.matches <- get.match_info(events.wc.18)
wc.22.matches <- get.match_info(events.wc.22)
wwc.19.matches <- get.match_info(events.wwc.19)
weuro.22.matches <- get.match_info(events.weuro.22)
```

Try to run only once:

```
# carry.wc.18 <- get.progressive(events.wc.18, "Carry")
# pPass.wc.18 <- get.progressive(events.wc.18, "Pass")
#
# carry.wc.22 <- get.progressive(events.wc.22, "Carry")
# pPass.wc.22 <- get.progressive(events.wc.22, "Pass")
#
# carry.wwc.19 <- get.progressive(events.wwc.19, "Carry")
# pPass.wwc.19 <- get.progressive(events.wwc.19, "Pass")
#
# carry.weuro.22 <- get.progressive(events.weuro.22, "Carry")
# pPass.weuro.22 <- get.progressive(events.weuro.22, "Pass")
```

```
carry.wc.18 <- read.csv("Data/Prog/wc18_carry")
pPass.wc.18 <- read.csv("Data/Prog/wc18_pPass")
```

```
carry.wc.22 <- read.csv("Data/Prog/wc22_carry")
pPass.wc.22 <- read.csv("Data/Prog/wc22_pPass")
```

```
carry.wwc.19 <- read.csv("Data/Prog/wwc19_carry")
pPass.wwc.19 <- read.csv("Data/Prog/wwc19_pPass")
```

```
carry.weuro.22 <- read.csv("Data/Prog/weuro22_carry")
pPass.weuro.22 <- read.csv("Data/Prog/weuro22_pPass")
```

All player stats should be normalized to per 90 minute. Players that have played less than 180 total minutes in the World Cup are excluded from rate statistics.

```
wc.22.ind <- compile.individualStats(events.wc.22,carry.wc.22,pPass.wc.22)
wc.18.ind <- compile.individualStats(events.wc.18,carry.wc.18,pPass.wc.18)

wwc.19.ind <- compile.individualStats(events.wwc.19,carry.wwc.19,pPass.wwc.19)
weuro.22.ind <- compile.individualStats(events.weuro.22,carry.weuro.22,pPass.weuro.22)

wc.18.ind.90 <- to.per90(wc.18.ind)
wc.22.ind.90 <- to.per90(wc.22.ind)

wwc.19.ind.90 <- to.per90(wwc.19.ind)
weuro.22.ind.90 <- to.per90(weuro.22.ind)
```

## Progressive Play

A progressive action is one that advances the ball 10 yards closer to goal than to furthest it has been in the last 6 relevant events in the sequence of play, or when the ball is entered into the 18 yard box. The common actions that can be considered progressive are passes (pP) and carries (pC).

```
summary(wc.22.ind.90 %>% arrange(desc(pC)) %>% select(team.name,pC,pP,X90s))
```

### Progressive summary of players at the 2022 World Cup

##	team.name	pC	pP	X90s
##	Length:294	Min. : 0.000	Min. : 0.800	Min. :2.011
##	Class :character	1st Qu.: 1.237	1st Qu.: 6.308	1st Qu.:2.681
##	Mode :character	Median : 2.263	Median :10.050	Median :3.000
##		Mean : 2.590	Mean :10.276	Mean :3.469
##		3rd Qu.: 3.460	3rd Qu.:13.659	3rd Qu.:4.000
##		Max. :11.875	Max. :24.639	Max. :7.667

```
summary(wc.18.ind.90 %>% arrange(desc(pC)) %>% select(pC,pP,X90s))
```

### Progressive summary of players at the 2018 World Cup

##	pC	pP	X90s
##	Min. : 0.000	Min. : 0.8571	Min. :2.022
##	1st Qu.: 1.000	1st Qu.: 6.7256	1st Qu.:2.883
##	Median : 2.290	Median : 9.8882	Median :3.000
##	Mean : 2.717	Mean :10.2675	Mean :3.619
##	3rd Qu.: 3.637	3rd Qu.:13.3333	3rd Qu.:4.017
##	Max. :13.400	Max. :30.3333	Max. :7.678

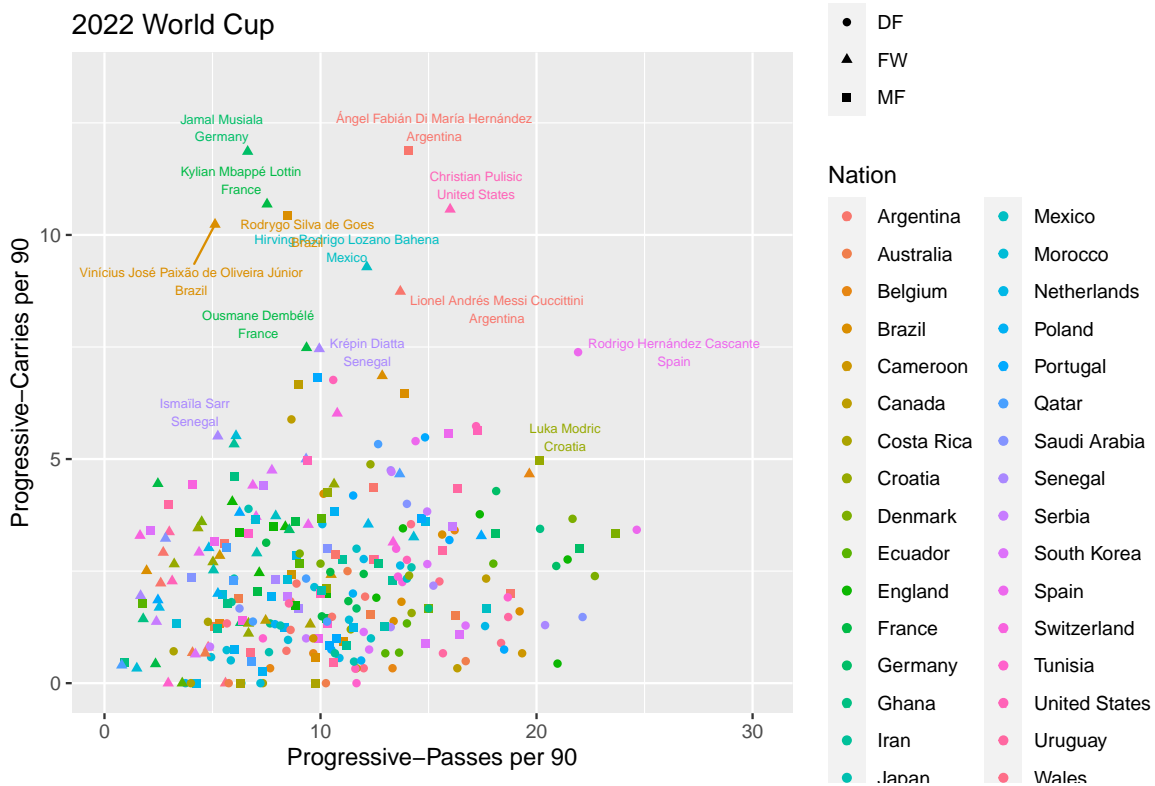
### Progression at the 2022 World Cup

```

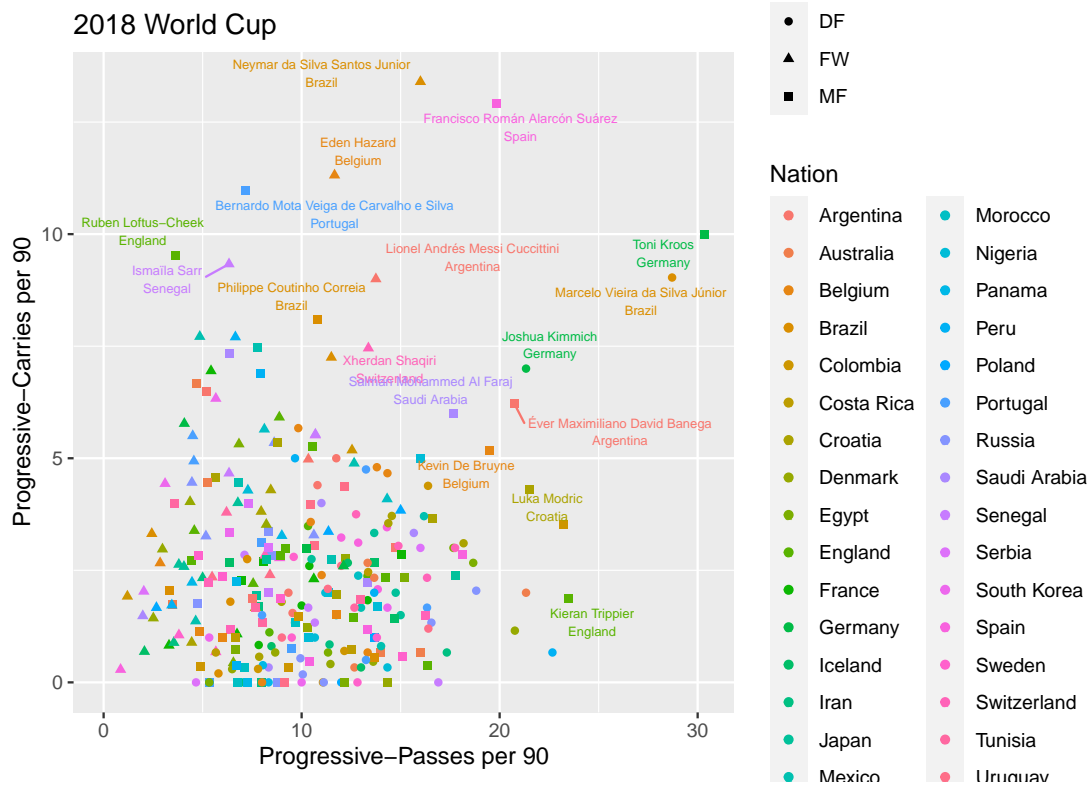
mPC <- max(wc.18.ind.90$pC, wc.22.ind.90$pC, wwc.19.ind.90$pC, weuro.22.ind.90$pC)
mPP <- max(wc.18.ind.90$pP, wc.22.ind.90$pP, wwc.19.ind.90$pP, weuro.22.ind.90$pP)

gplt <- ggplot(wc.22.ind.90,
  aes(y=pC,
      x=pP,
      # label=ifelse(
      #   labBool,
      #   paste(player.name,team.name,sep='\n'),
      #   NA
      # ),
      label=paste(player.name,team.name,sep='\n'),
      shape=factor(pos_group.name),
      color=factor(team.name)
  )
) +
  xlim(0, mPP) +
  ylim(0, mPC) +
  geom_point() +
  geom_text_repel(size=2) +
  labs(title = "2022 World Cup" +
  ylab("Progressive-Carries per 90" +
  xlab("Progressive-Passes per 90" +
  scale_colour_discrete(name="Nation") +
  scale_shape_discrete(name="Position") +
  coord_fixed(max(wc.22.ind.90$pP)/max(wc.22.ind.90$pC))
plot(gplt)

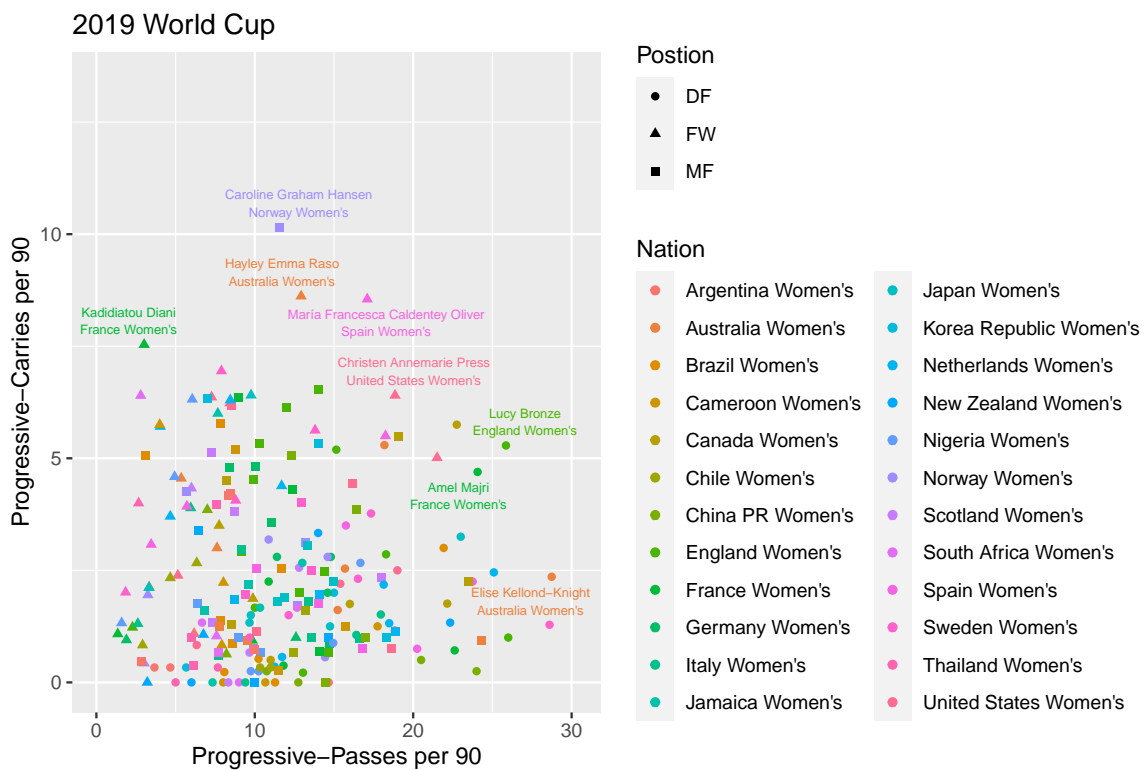
```



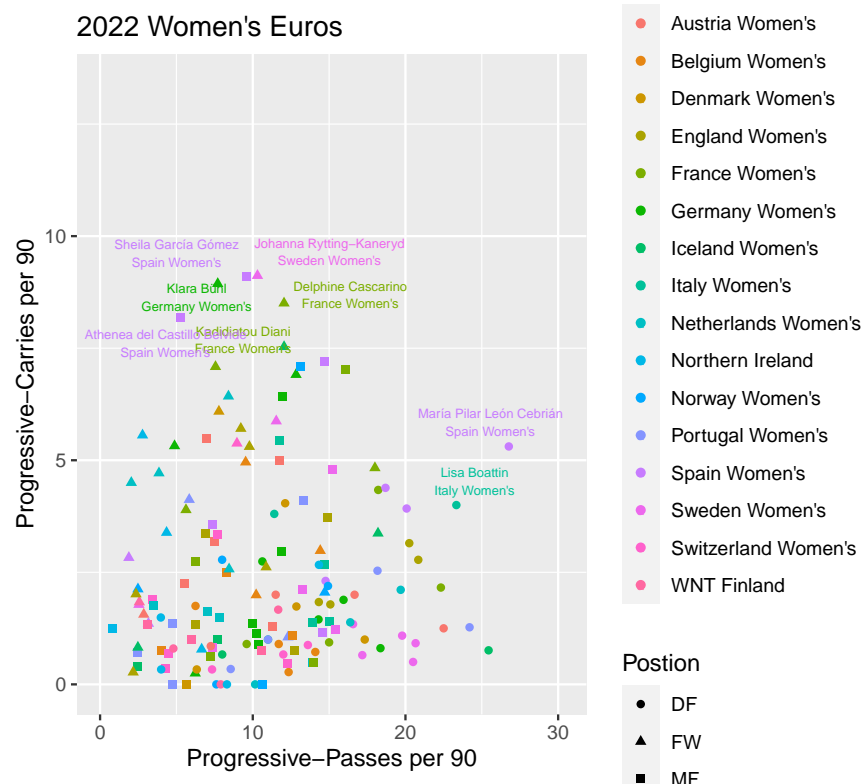
## Progression at the 2018 World Cup



## Progression at the 2019 World Cup



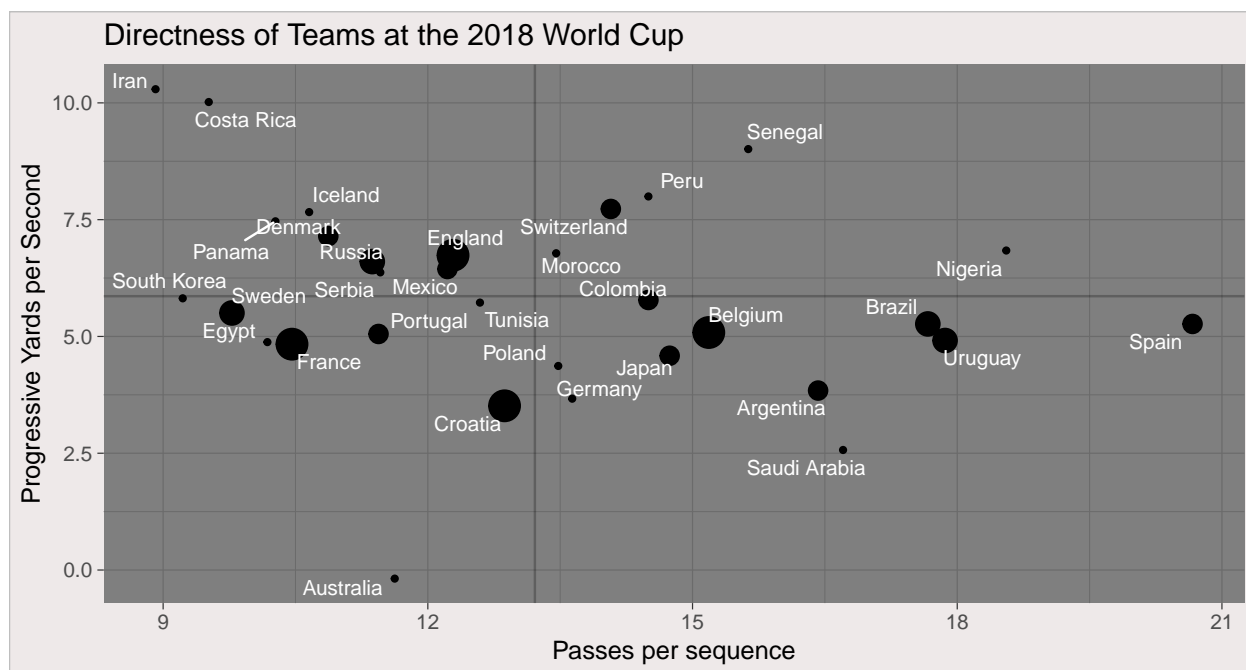
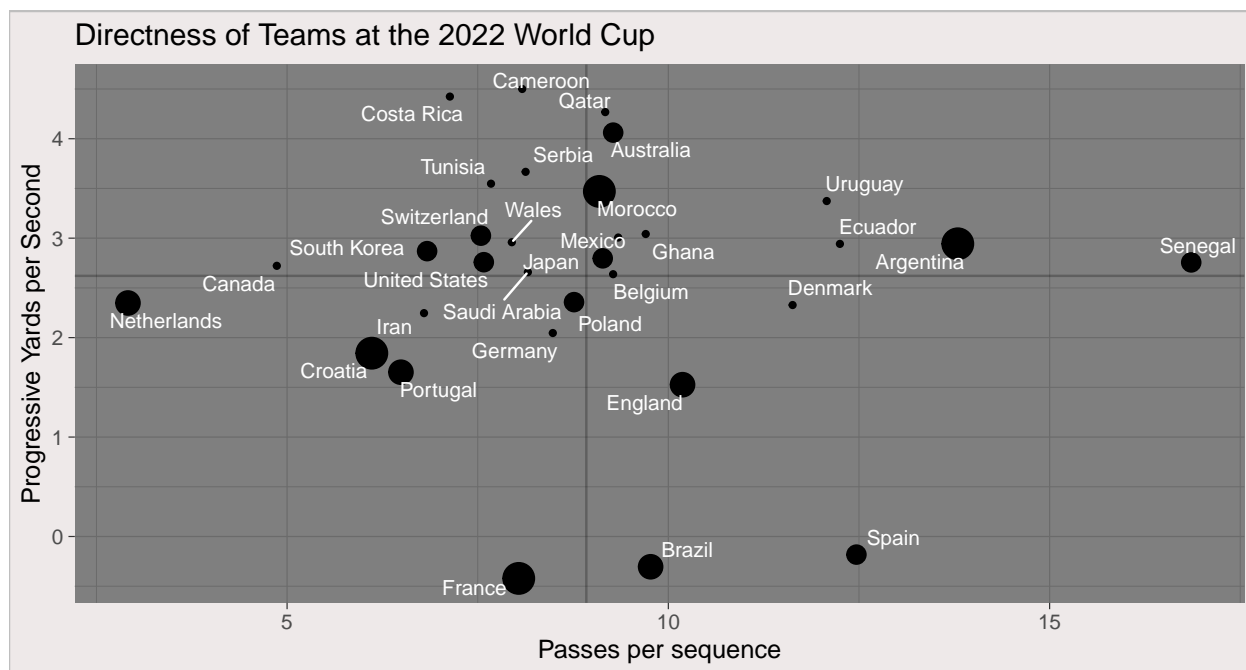
## Progression at the 2022 Women's Euros

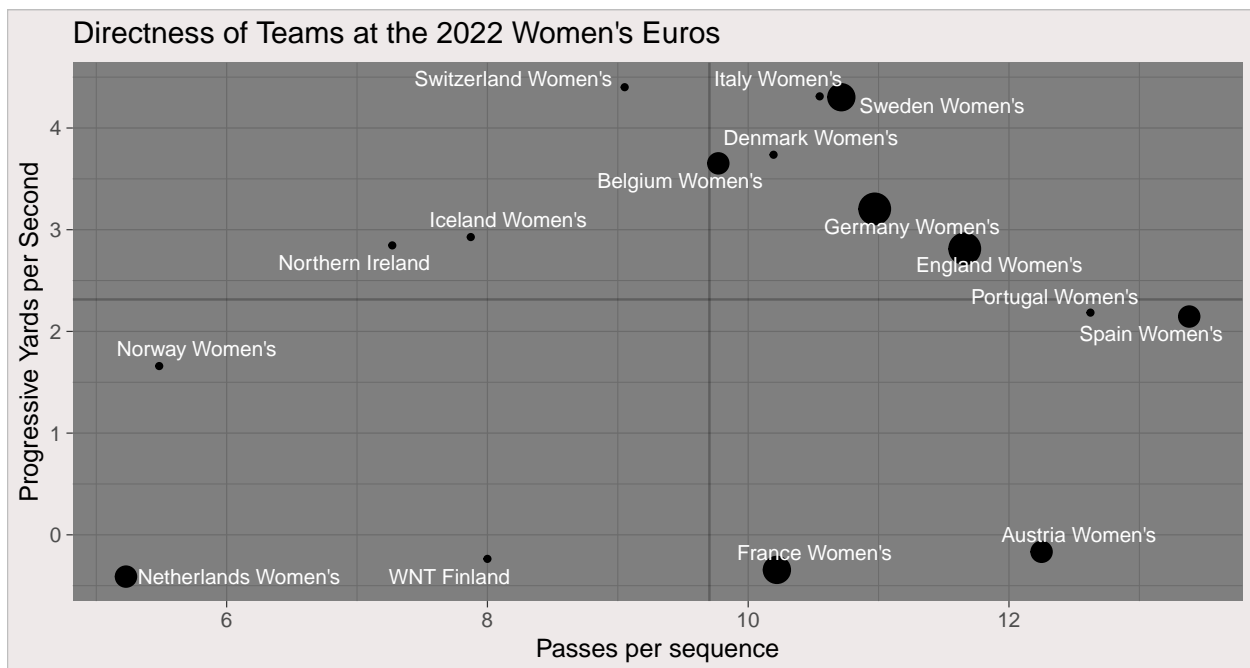
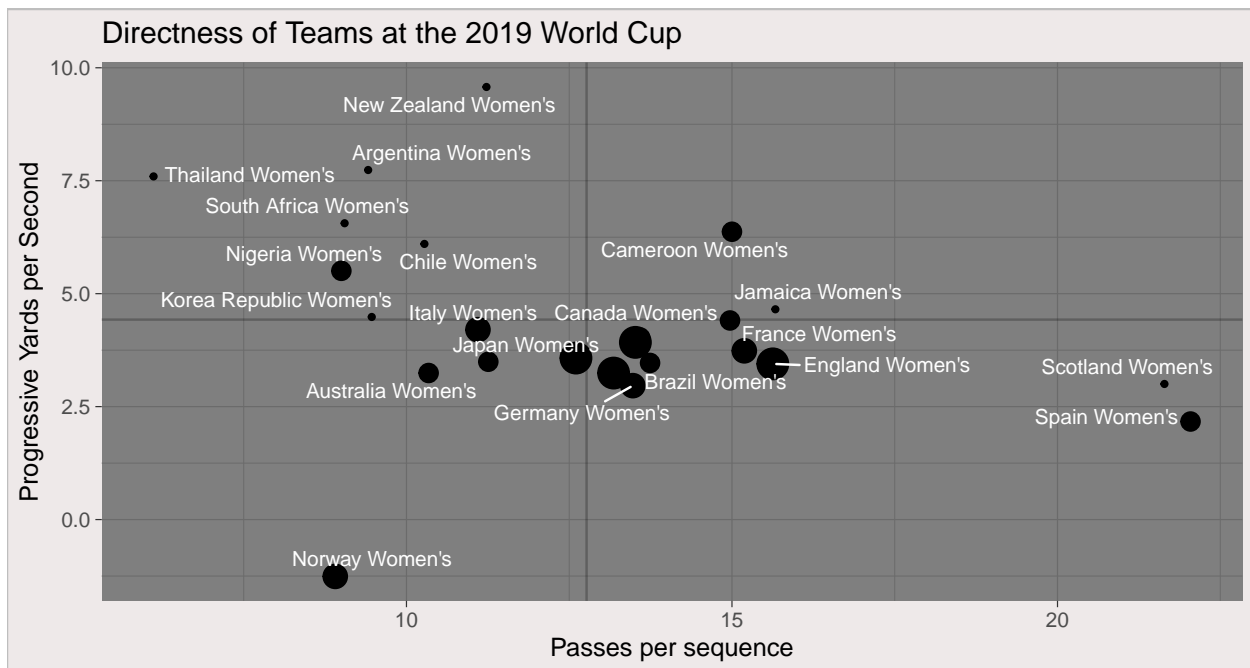


Fun little observation: Mapi Leon, the Spanish defender responsible for the most progressive passes and over 5 progressive carries per 90 at Euro22, is one of the exiled players from Jorge Vilda's World Cup squad. While the group of players has, she is one of the group refusing to play for the national team while Vilda remains its manager.

## Direct Team Play

Directness of play can be measured by the speed at which a team progresses the ball. The directness of a sequence play (a sequence being consecutive attacking actions, such as passes, carries, dribbles, shots) is the progressed distance from the first event in the sequence to the last per second (so divided by the duration of the sequence). The intricacy of team also relates to how direct they play, which is simply the number of passes per sequence. Generally less direct teams pass more, opting from to patiently move the ball backwards and side to side.





## Passing Networks

Passing networks are a fairly intuitive football-data visualization. The points/nodes represent each player on a given team, and the line/path connections between represent passes between two players. Nodes are placed at the average location that a given player passes the ball. Passes are normalized to per 90 minute rate. Connections are made when the first player passes to the second at least five times per 90. The number of passes per 90 scales the sizes of the nodes and lines. The network only represents the passes made by the first eleven, and it stops when the first sub is made for the team.

## 2022 World Cup Final

### Argentina vs. France

First sub in minute 63

3869685



The passing networks above can tell the story of a match (at least part of it) and the roles of certain players. First looking at Argentina's network, nearly every players has at least three line connected to them; this shows that the team was strongly connected by passing and passing well. Additionally we can see that on a simple level, Argentina setup in a 4-3-3 with the fullbacks pushing up slightly further and Mac Allister playing more advanced than his midfield partners. More interestingly, there is an obvious difference in how their outside forwards play. Di Maria on the left is pushed high and wide, reflecting how his play focused around taking on defenders in wide-attacking areas to create chances, such as winning a penalty for the first goal. On the right is Messi. Throughout Messi's entire career, his positioning has been basically the same with some variations on it: He plays toward the right, and comes more centrally and deeper to get the ball. His positioning for Argentina follows this rule, with his average passing position being nearly as deep as the highest midfielder, and his width being on the border of the central area and right half-space.



## France vs. Argentina

First sub in minute 40

3869685



France's passing network tells a completely different story. As aside from the defense and the outside-backs to the wingers, there is very little connecting. Perhaps most concerning, the midfield has almost no connections at all. Tchouameni has no connections with his defense, showing him completely unable to connect the defense and attack. Meanwhile, Greizman is on island high up the pitch. The network also shows only one link to the team's centerforward (Giroud), which is a faint link from his goalkeeper. If the lack of passing and connectedness doesn't show that France struggled to generate attacks, the fact that their striker couldn't get the ball should.

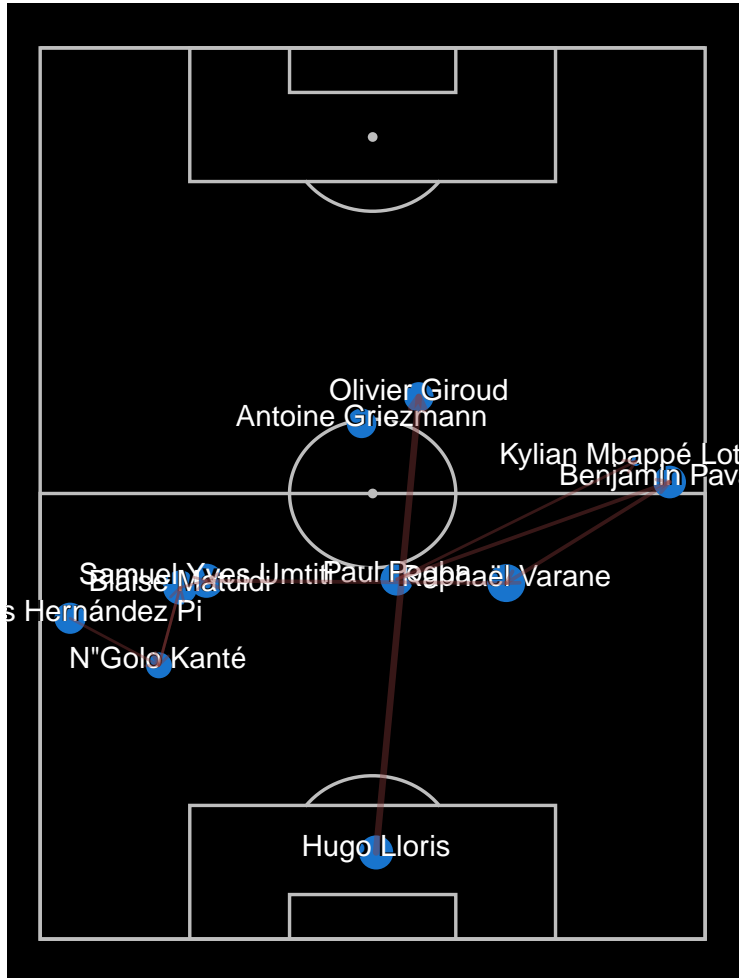
Watching the match backs up these conclusions, with France looking so lost that they made substitutions before halftime, and Argentina completely controlling the game until Di Maria's departure in the 62 minute. What happened after that was a different story, but each network cut off before the game flipped.

## 2018 World Cup Final

France vs. Croatia

First sub in minute 54

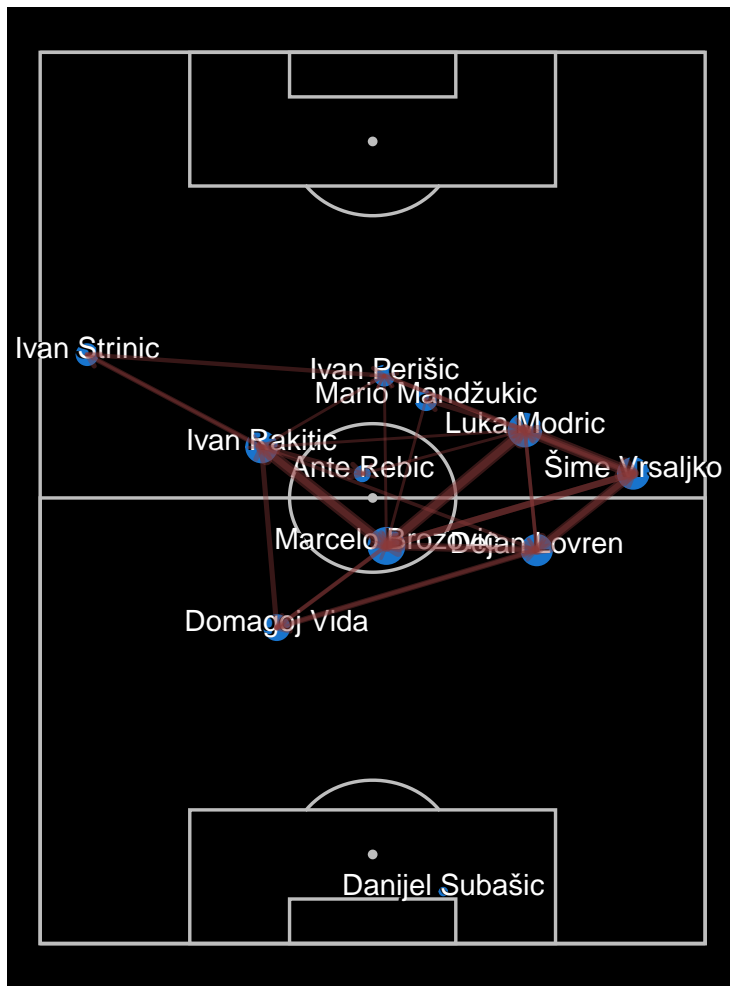
8658



## Croatia vs. France

First sub in minute 70

8658



## Defense

### Ball Recoveries

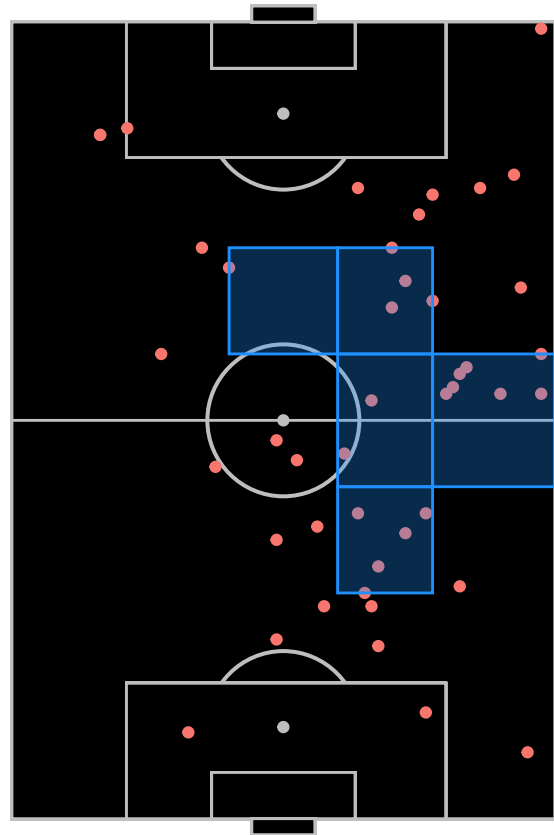
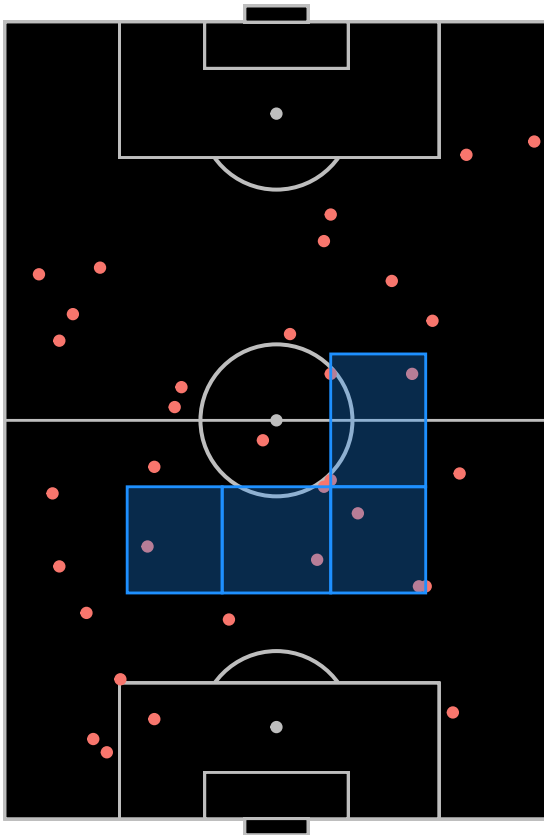
#### World Cup 2018

Here are some ball recovery maps from the 2018 world cup. These are 4 of the top 5 outfield players in totals in that category. Which is a great excuse to behold the beauty of an N'Golo Kante Ball Recovery Map.

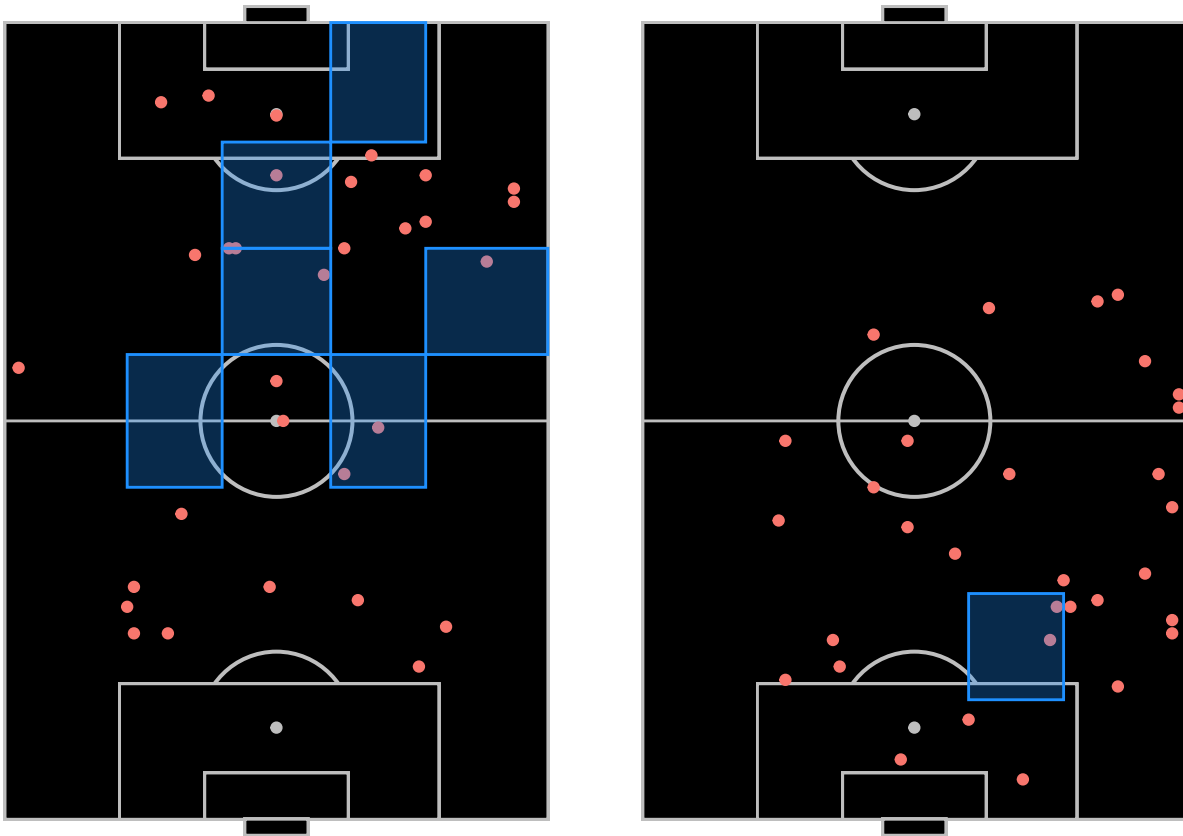
```
## # A tibble: 560 x 4
## # Groups:   player.id, player.name [560]
##   player.id player.name     team.name total_br
##   <int> <chr>             <chr>      <dbl>
## 1    3509 Thibaut Courtois Belgium      44
## 2    5463 Luka Modrić      Croatia      44
## 3    3468 Jordan Pickford  England      39
## 4    3444 Danijel Subašić    Croatia      38
```

##	5	5171	Roman Zobnin	Russia	35
##	6	3961	N'Golo Kanté	France	34
##	7	3089	Kevin De Bruyne	Belgium	31
##	8	5485	Raphaël Varane	France	31
##	9	5613	Viktor Claesson	Sweden	31
##	10	5466	Oghenekaro Etebo	Nigeria	30
##	#	i	550	more rows	

N'Golo Kanté, 34 Successful Ball Recoveries    Luka Modric, 44 Successful Ball Recoveries



Kevin De Bruyne, 31 Successful Ball Re Raphaël Varane, 31 Successful Ball Re



## PPDA

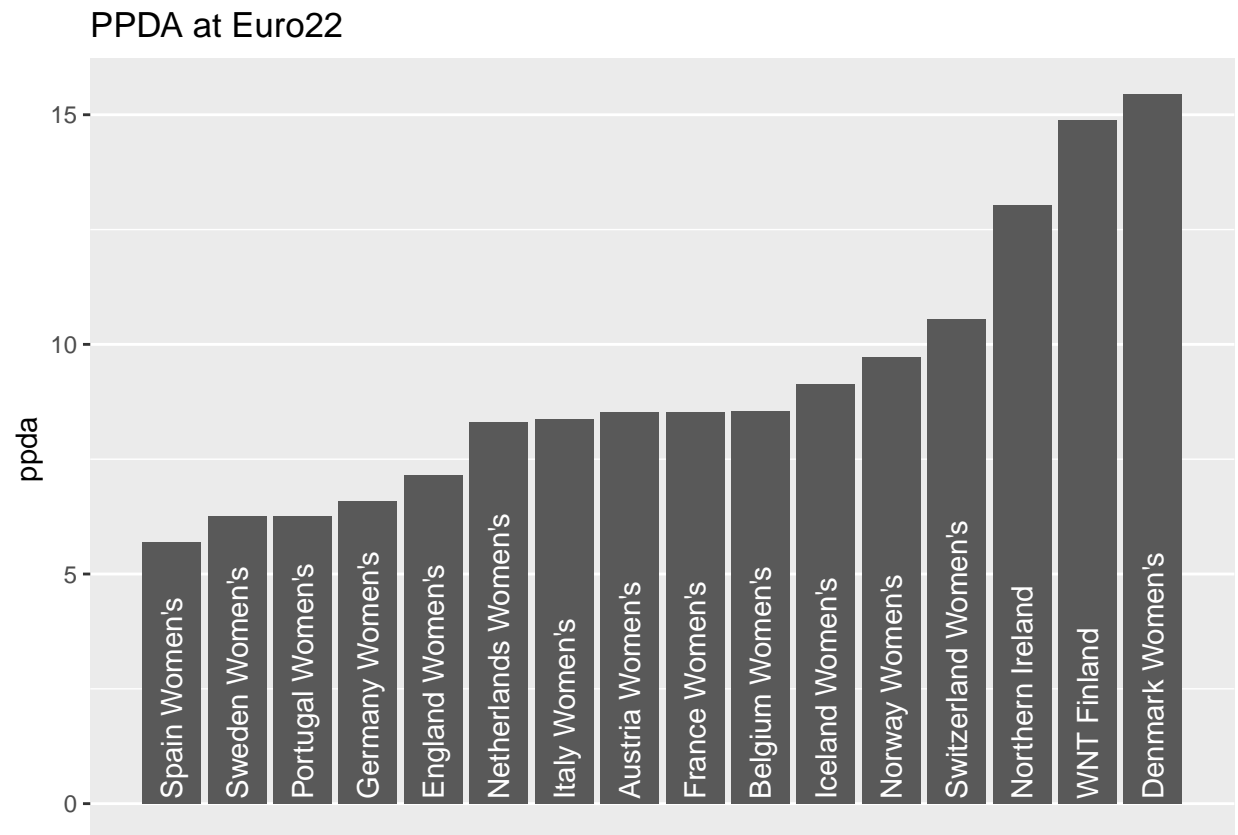
PPDA (Passes Per Defensive Action) is probably the leading statistic in measuring a teams the intensity and effectiveness of a teams high press. Looking only at the attacking 60% of the pitch, PPDA is calculated by dividing the total passes of the opposition by the total number of defensive actions (tackle attempts, interceptions, and fouls committed) perform. Lower the PPDA that teams emphasize pressuring there opposition with urgency high up the pitch. Higher PPDA shows that a team is content sitting back and on letting the opposition come to them.

```
ppda.all <- c()
for (tm.id in unique(events.weuro.22$team.id)) {ppda.all<-c(ppda.all,get.PPDA(events.weuro.22,tm.id))}
ppda.teams <- merge(
  unique(events.weuro.22[,c("team.name","team.id")]),
  data.frame(team.id=unique(events.weuro.22$team.id), ppda=ppda.all),
  by="team.id"
) %>% arrange(ppda)
gplt <- ggplot(ppda.teams, aes(y=ppda, x=1:length(ppda.all), label=team.name)) +
  geom_col() +
  geom_text(angle=90, hjust="left", y=0.1, color="white") +
  theme(
    axis.title.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank(),
```

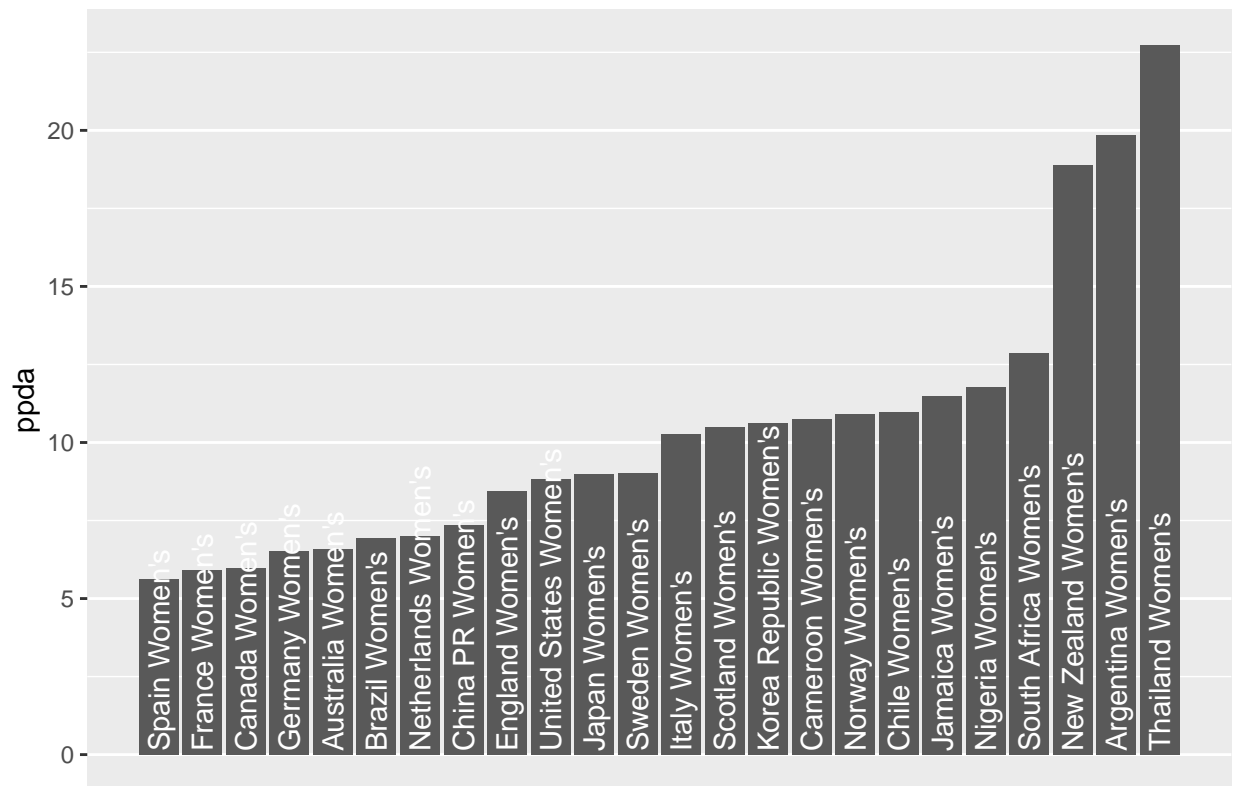
```

panel.grid.major.x = element_blank(),
panel.grid.minor.x = element_blank()
) +
labs(title = "PPDA at Euro22")
plot(gplt)

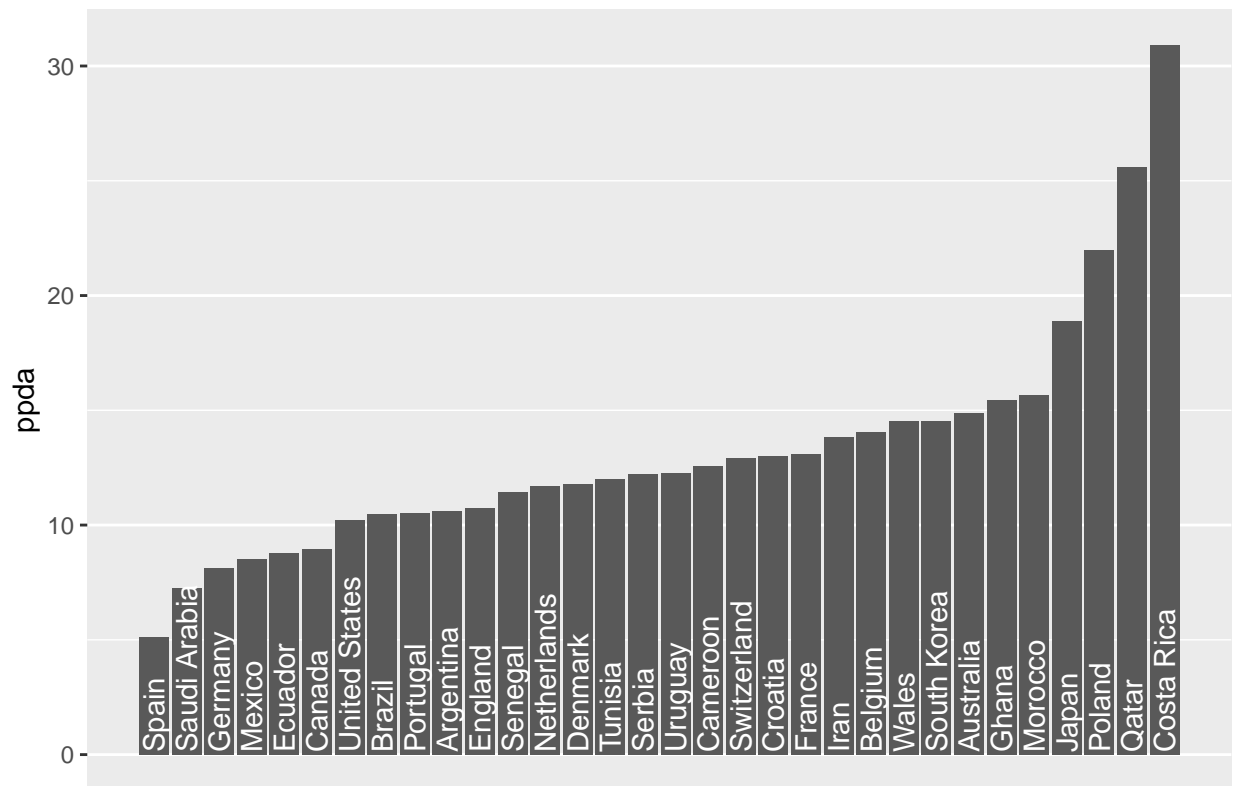
```



## PPDA at World Cup 2019



## PPDA at World Cup 2022





## PPDA at World Cup 2018

