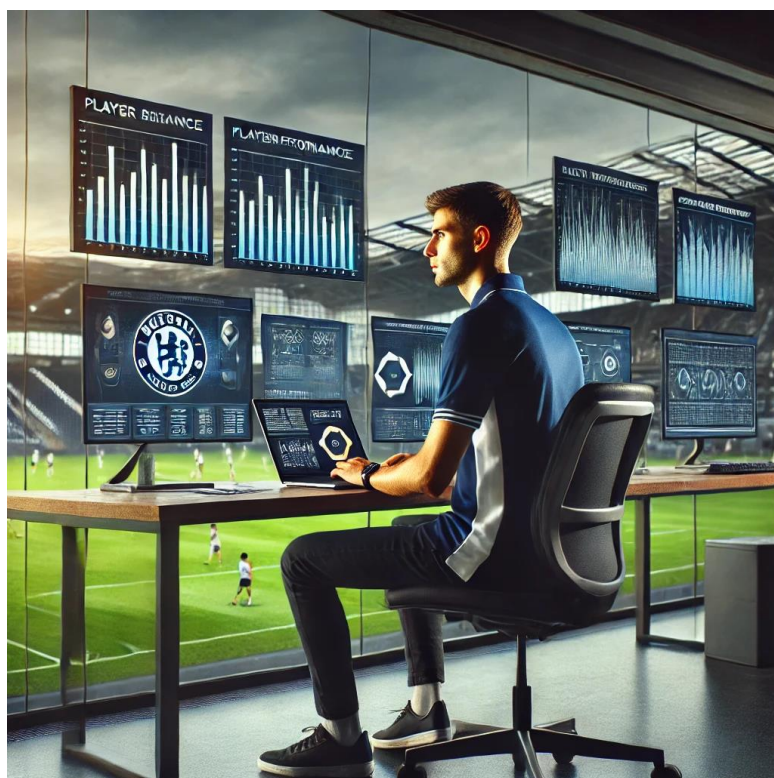


Master em Big Data Aplicado ao Futebol

Módulo 4: Linguagem de Programação R



Exercício Prático

Trabalho realizado por: Gonalo Mendes Rodrigues

Desafio

O desafio consiste em colocar um analista de dados no centro das decisões estratégicas de um clube de futebol, utilizando **R** para analisar e visualizar dados de desempenho. A tarefa inclui avaliar jogadores individualmente, como **Leandro Paredes**, **Rodri**, **Messi** e **Sterling**, através de mapas de ações defensivas, mapas de calor e comparações de dribles. Além disso, envolve uma análise coletiva do **posicionamento médio**, **mapa de passes** e **golos esperados (xG)** das equipas, fornecendo insights para otimizar a tática em campo.

O objetivo final é produzir um **relatório detalhado** e um **script em R bem documentado**, organizando os outputs num **arquivo comprimido** pronto para apresentação à equipa técnica. A análise servirá como base para melhorar o desempenho da equipa e criar estratégias vencedoras.

1) Leandro Paredes (PSG): O treinador quer uma visão completa das ações defensivas deste jogador. Consegue cobrir as áreas mais perigosas? Está a proteger eficazmente o setor defensivo?
Tarefa: Criar a representação das ações defensivas de Leandro Paredes no relvado.

De modo a responder a este desafio, comecei por analisar quais eram as dimensões do campo dos dados que nos foram fornecidos para que a posição das ações dos jogadores fosse representada corretamente nos gráficos.

```
#Obter dimensões do campo  
max_pos_x <- max(dados$pos_x, na.rm = TRUE)  
max_pos_y <- max(dados$pos_y, na.rm = TRUE)
```

De seguida criei um campo com dimensões customizadas para utilizar na resposta aos vários exercícios:

```
#criar campo com dimensões adequadas aos dados que temos
pitch_custom <- list(
  length = 105,
  width = 68,
  penalty_box_length = 16.5,
  penalty_box_width = 40.3,
  six_yard_box_length = 5.5,
  six_yard_box_width = 18.32,
  penalty_spot_distance = 11,
  goal_width = 7.32,|
  origin_x = 0,
  origin_y = 0
)
```

O passo seguinte consistiu em filtrar os dados do Paredes e as suas ações defensivas:

```
# Filtrar os dados do Paredes
dados_paredes <- dados %>% filter(code == "8. Paredes")

# Ações defensivas de interesse
acoes_desejadas <- c(
  "Interceptions",
  "Challenges (won)",
  "Tackles (Successful actions)",
  "Picking-ups",
  "Challenges (lost)",
  "Tackles (Unsuccessful actions)",
  "Fouls"
)
```

De modo a tornar a interpretação da performance do Paredes mais simples decidi agrupar as diferentes ações defensivas em 4 categorias:

```
# Criar uma nova coluna para classificar as ações
dados_defensivos_paredes <- dados_defensivos_paredes %>%
  mutate(
    Categoria = case_when(
      Action == "Fouls" ~ "Falta",
      Action == "Interceptions" ~ "Intercepção",
      Action %in% c("Challenges (won)", "Tackles (Successful actions)", "Picking-ups") ~ "Ação Defensiva Bem-sucedida",
      Action %in% c("Challenges (lost)", "Tackles (Unsuccessful actions)") ~ "Ação Defensiva Mal-sucedida"
    )
  )
```

Para complementar a informação que vou utilizar no gráfico decidi calcular os totais e percentagens das ações defensivas do Paredes durante o jogo:

```
# Calcular valores totais e percentagens
total_pickups <- sum(dados_defensivos_paredes$Action == "Picking-ups", na.rm = TRUE)
tackles_ganhos <- sum(dados_defensivos_paredes$Action == "Tackles (Successful actions)", na.rm = TRUE)
tackles_totais <- tackles_ganhos + sum(dados_defensivos_paredes$Action == "Tackles (Unsuccessful actions)", na.rm = TRUE)
challenges_ganhos <- sum(dados_defensivos_paredes$Action == "Challenges (won)", na.rm = TRUE)
challenges_totais <- challenges_ganhos + sum(dados_defensivos_paredes$Action == "Challenges (lost)", na.rm = TRUE)

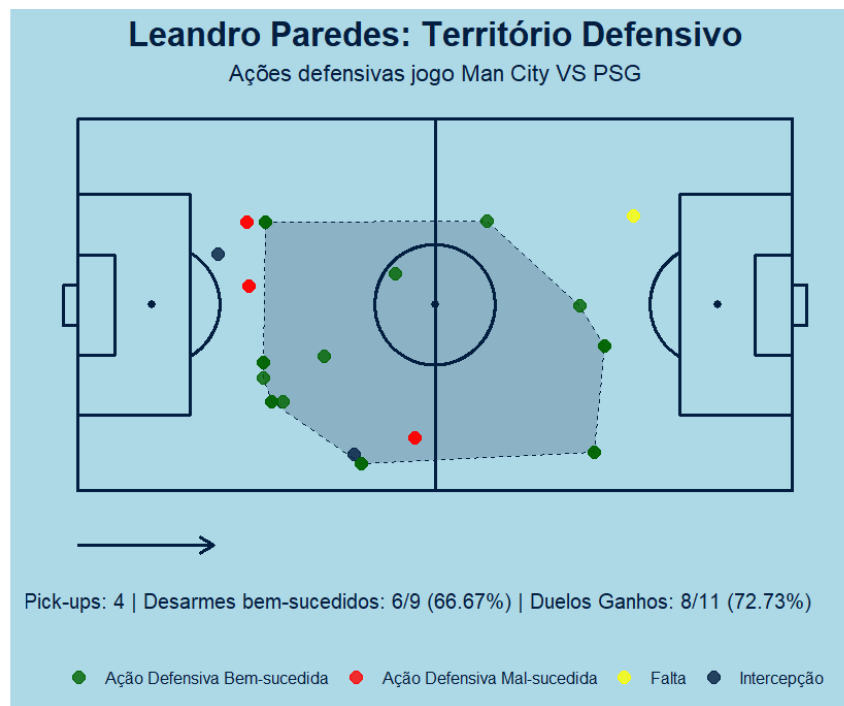
percent_tackles <- ifelse(tackles_totais > 0, round((tackles_ganhos / tackles_totais) * 100, 2), 0)
percent_challenges <- ifelse(challenges_totais > 0, round((challenges_ganhos / challenges_totais) * 100, 2), 0)
```

Calculei uma Convex Hull com base nas ações defensivas bem sucedidas do Paredes

para representar a área que ele conseguiu defender de forma efetiva:

```
# Calcular a convex hull das ações bem-sucedidas  
bem_sucedidas <- dados_defensivos_paredes %>%  
  filter(Categoria == "Ação Defensiva Bem-sucedida")  
hull_indices <- chull(bem_sucedidas$pos_x, bem_sucedidas$pos_y)  
hull_points <- bem_sucedidas[hull_indices, ]
```

De seguida com recurso ao ggplot elaborei o gráfico pretendido:



O gráfico mostra que Leandro Paredes tem uma boa cobertura defensiva no meio-campo e zonas próximas da área, destacando uma taxa de sucesso de 66.67% nos desarmes e 72.73% nos duelos. No entanto, também evidencia algumas ações defensivas mal-sucedidas, sugerindo possíveis fragilidades em certos momentos do jogo.

2) Rodri (Man. City): Um pilar no meio-campo. O objetivo é entender os seus movimentos em campo através de mapas de percurso e calor, destacando as zonas de maior atividade.

Tarefa: Mostrar o percurso e o mapa de calor de Rodri.

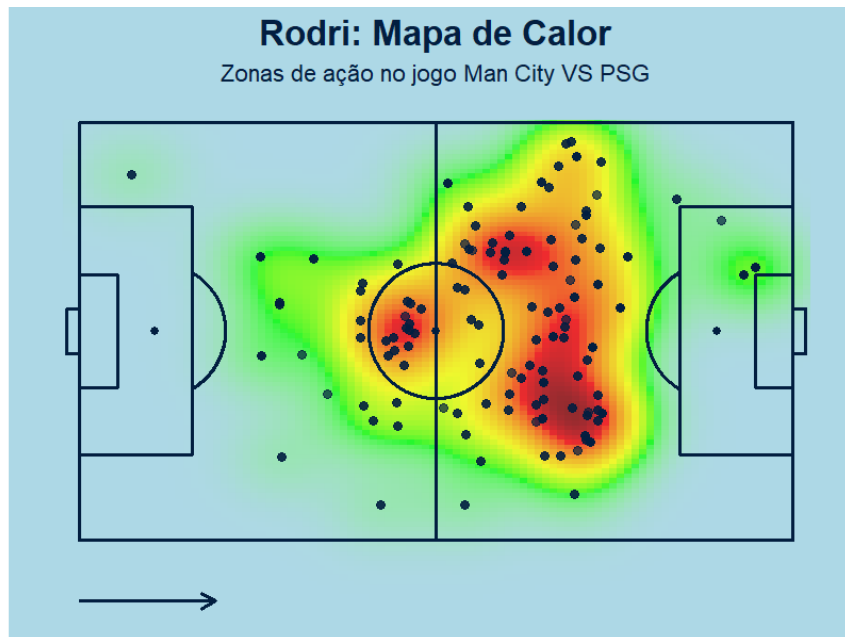
Comecei por filtrar os dados do Rodri:

```
# Filtrar os dados do Rodri
dados_rodri <- dados %>% filter(code == "16. Rodri")
```

Depois, com recurso ao ggplot criei um mapa de calor das ações do Rodri no jogo Man City vs PSG, destacando as zonas onde esteve mais ativo no campo. O `stat_density_2d` gera o mapa de calor com uma escala de cores progressiva de verde a vermelho escuro, que são as zonas de maior atividade. O campo é desenhado com `annotate_pitch` como no exercício 1, enquanto `geom_point` adiciona os pontos exatos das ações do jogador.

```
# ---- MAPA DE CALOR ---- #
ggplot() +
  # Heatmap
  stat_density_2d(
    data = dados_rodri,
    aes(x = pos_x, y = pos_y, fill = ..density..),
    geom = "raster", contour = FALSE, alpha = 0.8
  ) +
  # escala de cores
  scale_fill_gradientn(
    colours = c("lightblue", "green", "yellow", "orange", "red", "darkred"),
    guide = "none"
  ) +
  annotate_pitch(
    dimensions = pitch_custom,
    colour = "#021e40", linewidth = 0.8, fill = NA, limits = FALSE) +
  annotate(
    "segment",
    x = 0, xend = 20,
    y = -10, yend = -10,
    arrow = arrow(length = unit(0.3, "cm")),
    colour = "#021e40", linewidth = 0.8
  ) +
  # Pontos das ações do Rodri
  geom_point(
    data = dados_rodri,
    aes(x = pos_x, y = pos_y),
    size = 2, alpha = 0.7, colour = "#021e40"
  ) +
  labs(
    title = "Rodri: Mapa de Calor",
    subtitle = "Zonas de ação no jogo Man City VS PSG",
    x = NULL,
    y = NULL
  ) +
  theme_pitch() +
  theme(
    plot.title = element_text(face = "bold", size = 18, hjust = 0.5, color = "#021e40"),
    plot.subtitle = element_text(size = 12, hjust = 0.5, color = "#021e40"),
    plot.background = element_rect(fill = "lightblue", color = NA), # Fundo azul escuro
    panel.background = element_rect(fill = "lightblue", color = NA)
  )
```

O mapa de calor apresenta as zonas de maior atividade do Rodri durante o jogo Man City vs PSG, evidenciando a sua influência no meio-campo defensivo e central. As áreas de maior intensidade, representadas a vermelho escuro, indicam os locais onde esteve mais presente, e demonstram a sua relevância na construção de jogo e na recuperação da posse de bola. Observa-se que Rodri cobre uma vasta área do meio-campo, com menor incidência nas alas e na grande área adversária, o que confirma o seu papel como elemento fundamental na transição defensiva e na organização do jogo da equipa.



3) Leo Messi (PSG) e Sterling (Man. City): Estes dois jogadores são conhecidos pelos seus dribles, mas qual deles é mais eficaz? Vais comparar dribles certos e errados, apresentando conclusões claras à equipa técnica.

Tarefa: Representar os dribles bem-sucedidos e falhados de Leo Messi e Sterling.

Neste exercício, comecei por filtrar os dados dos dribles de ambos os jogadores:

```
# Filtrar os dados do Messi e Sterling
dados_messi <- dados %>% filter(code == "30. Messi")

dados_sterling <- dados %>% filter(code == "7. Sterling")

# Filtrar dados de drible de Messi e Sterling
dados_messi_dribbles <- dados_messi %>%
  filter(Action %in% c("Dribbles (Successful actions)", "Dribbles (Unsuccessful actions)"))

dados_sterling_dribbles <- dados_sterling %>%
  filter(Action %in% c("Dribbles (Successful actions)", "Dribbles (Unsuccessful actions)"))
```

O passo seguinte consistiu no cálculo do número total de dribles de cada jogador e a percentagem de dribles corretamente efetuados:

```
#Calcular valores totais e percentagens
messi_total <- nrow(dados_messi_dribbles)
messi_sucess <- nrow(dados_messi_dribbles %>% filter(Action == "Dribbles (Successful actions)"))
messi_percent <- round((messi_sucess / messi_total) * 100, 2)

sterling_total <- nrow(dados_sterling_dribbles)
sterling_sucess <- nrow(dados_sterling_dribbles %>% filter(Action == "Dribbles (Successful actions)"))
sterling_percent <- round((sterling_sucess / sterling_total) * 100, 2)
```

Procedi à criação de um gráfico de barras para visualizar a percentagem de dribles eficazes de cada jogador:

```

# Criar o gráfico de barras
grafico_barras <- ggplot(data = data.frame(
  Player = c("Messi", "Sterling"),
  Percent = c(messi_percent, sterling_percent)
), aes(x = Player, y = Percent, fill = Player)) +
  geom_bar(stat = "identity", color = "black", alpha = 0.8, width = 0.6) +
  scale_fill_manual(values = c("Messi" = "#021e40", "Sterling" = "lightblue")) +
  labs(
    x = NULL,
    y = "% de Dribles bem sucedidos"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color = "#021e40"),
    axis.title.y = element_text(color = "#021e40", size = 12),
    axis.text = element_text(color = "#021e40"),
    plot.background = element_rect(fill = "lightblue", color = NA),
    panel.background = element_rect(fill = "lightblue", color = NA),
    legend.position = "none",
    panel.grid.major.x = element_blank()
  )

```

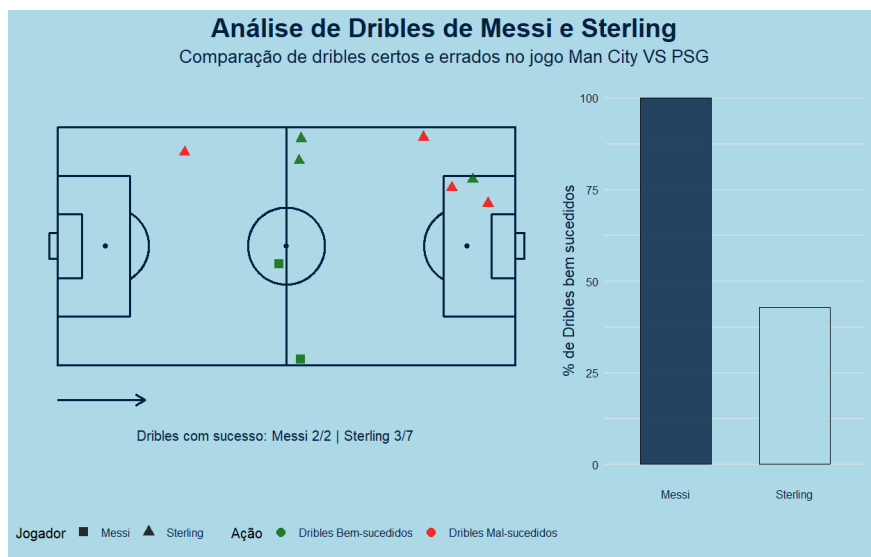
Para complementar a análise representei no campo os vários dribles de Messi e Sterling que nos permite verificar onde foram efetuados os dribles de ambos e se foram bem ou mal sucedidos:

```

# Criar o campo com dribles representados
campo <- ggplot() +
  annotate_pitch(
    dimensions = pitch_custom,
    colour = "#021e40", linewidth = 0.8, fill = "lightblue", limits = FALSE) +
  geom_point(
    data = dados_dribbles,
    aes(x = pos_x, y = pos_y, color = Action, shape = Player),
    size = 3, alpha = 0.8
  ) +
  scale_color_manual(
    values = c(
      "Dribbles (Successful actions)" = "darkgreen",
      "Dribbles (Unsuccessful actions)" = "red"
    ),
    labels = c(
      "Dribbles (Successful actions)" = "Dribles Bem-sucedidos",
      "Dribbles (Unsuccessful actions)" = "Dribles Mal-sucedidos"
    ),
    name = "Ação"
  ) +
  scale_shape_manual(
    values = c("Messi" = 15, "Sterling" = 17),
    name = "Jogador"
  ) +
  # Adicionar a legenda personalizada abaixo do campo
  annotate(
    "text",
    x = 50, y = -20,
    label = paste0(
      "Dribbles com sucesso: Messi ", messi_sucess, "/", messi_total,
      " | Sterling ", sterling_sucess, "/", sterling_total
    ),
    hjust = 0.5, size = 3.9, color = "#021e40"
  ) +
  annotate(
    "segment",
    x = 0, xend = 20,
    y = -10, yend = -10,
    arrow = arrow(length = unit(0.3, "cm")),
    colour = "#021e40", linewidth = 0.8
  ) +
  labs(
    #title = "Eficácia de Drible de Messi e Sterling",
    #subtitle = "Comparação de dribles certos e errados no jogo Man City VS PSG",
    x = NULL,
    y = NULL
  ) +
  theme_pitch() +
  theme(

```

No final combinei os dois gráficos:



Observa-se que Messi realizou dois dribles e teve uma taxa de sucesso de 100%, enquanto Sterling tentou sete dribles, mas conseguiu completar apenas três, resultando numa taxa de sucesso inferior a 50%. O gráfico de barras reforça essa diferença, demonstrando a superior eficácia de Messi nas situações de um contra um. A distribuição dos dribles no campo também indica que Sterling tentou mais vezes romper a defesa adversária, mas com menor eficiência, o que pode sugerir maior risco ou dificuldades em ultrapassar marcadores.

4) Posicionamento Médio: O treinador quer ver o posicionamento médio dos jogadores titulares de dois grandes clubes. Ele espera que inclua o ConvexHull para destacar a ocupação de espaço, idealmente num gráfico que apresente as duas equipas lado a lado. Tarefa: Mostrar o posicionamento médio dos titulares das duas equipas.

No 4º exercício, a primeira coisa a fazer é filtrar os dados pelos jogadores titulares:


```
#Filtrar os jogadores titulares
jogadores_titulaes <- c(
  "31. Ederson",
  "27. Cancelo",
  "3. R. Dias",
  "5. Stones",
  "2. Walker",
  "11. Zinchenko",
  "16. Rodri",
  "8. Gundogan",
  "7. Sterling",
  "20. B. Silva",
  "26. Mahrez",
  "1. K. Navas",
  "25. N. Tavares",
  "3. Kimpembe",
  "5. Marquinhos",
  "2. Hakimi",
  "27. Gueye",
  "8. Paredes",
  "21. A. Herrera",
  "10. Neymar",
  "7. Mbappe",
  "30. Messi"
)

dados_titulares <- dados %>% filter(code %in% jogadores_titulaes)
```

Depois calculei a posição média de cada jogador:

```
dados_titulares_medio <- dados_titulares %>%
  group_by(Team, code) %>%
  summarise(
    pos_x_medio = mean(pos_x, na.rm = TRUE),
    pos_y_medio = mean(pos_y, na.rm = TRUE)
  )
```

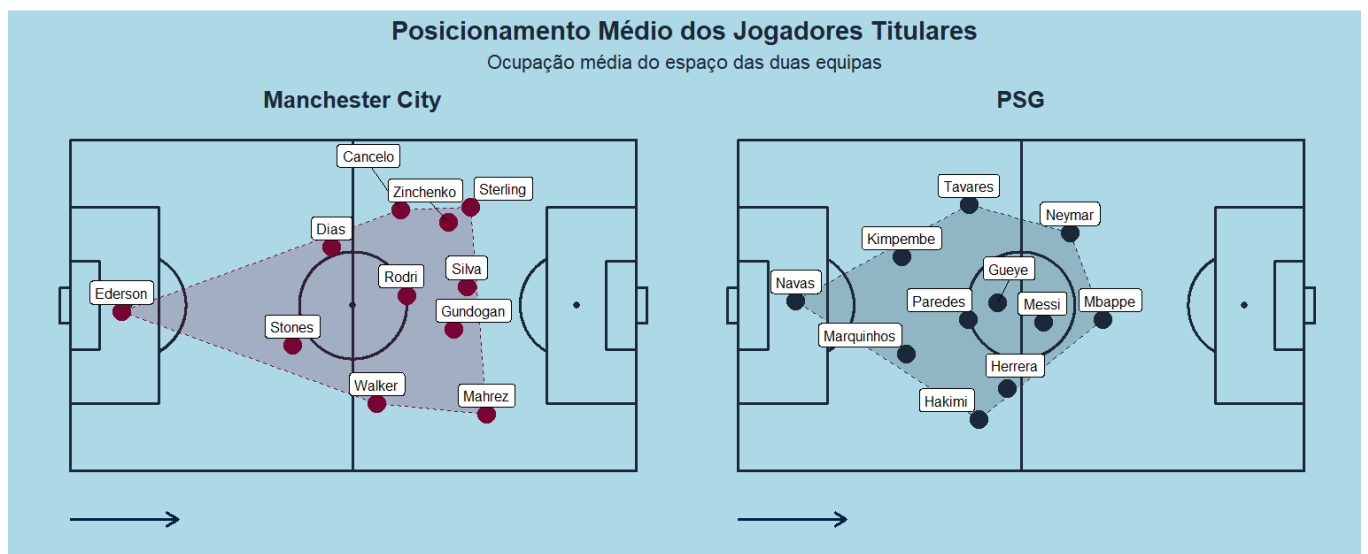
De seguida, com base no posicionamento dos jogadores de cada equipa, calculei a Convex hull:

```
# Adicionar o Convex Hull por equipa
convex_hull_data <- dados_titulares_medio %>%
  group_by(Team) %>%
  slice(chull(pos_x_medio, pos_y_medio))
```

Elaborei o gráfico para cada equipa em separado com a função `geom_polygon` a permitir desenhar a Convex Hull e o `geom_point` a permitir representar a posição média de cada jogador. A biblioteca `ggrepel` permite que os nomes dos jogadores não fiquem sobrepostos uns sobre os outros na elaboração do gráfico.

```
# Gráfico para Manchester City
grafico_team1 <- ggplot() +
  annotate_pitch(dimensions = pitch_custom, colour = "#1B2838", linewidth = 0.8, fill = "lightblue", limits = FALSE) +
  geom_polygon(
    data = convex_hull_data %>% filter(Team == team1),
    aes(x = pos_x_medio, y = pos_y_medio),
    fill = cores_equipas[team1], alpha = 0.2, colour = cores_equipas[team1], linetype = "dashed"
  ) +
  geom_point(
    data = dados_titulares_medio %>% filter(Team == team1),
    aes(x = pos_x_medio, y = pos_y_medio),
    size = 5, colour = cores_equipas[team1]
  ) +
  geom_label_repel(
    data = dados_titulares_medio %>% filter(Team == team1),
    aes(x = pos_x_medio, y = pos_y_medio, label = name),
    fill = "white", color = "black", size = 3, label.size = 0.3,
    nudge_y = 3, # Pequeno deslocamento para evitar sobreposição
    force = 2, # Aumenta a força da separação
    max.overlaps = Inf
  ) +
  annotate(
    "segment",
    x = 0, xend = 20,
    y = -10, yend = -10,
    arrow = arrow(length = unit(0.3, "cm")), |
    colour = "#021e40", linewidth = 0.8
  ) +
  labs(title = team1, x = NULL, y = NULL) +
  theme_pitch() +
  theme(
    panel.grid = element_blank(),
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color = "#1B2838"),
    plot.background = element_rect(fill = "lightblue", color = NA),
    panel.background = element_rect(fill = "lightblue", color = NA)
  )
}
```

No final combinei o gráfico de ambas as equipas:



O Manchester City adota uma estrutura mais alargada, com uma ocupação mais equilibrada do campo e uma defesa posicionada de forma ampla para facilitar a saída de bola. O meio-campo, com Rodri, Silva e Gundogan, está bem distribuído e próximo dos extremos (Sterling e Mahrez), o que sugere uma abordagem de posse e circulação de bola.

O PSG, por outro lado, apresenta um bloco mais compacto, especialmente no meio-campo, com Paredes, Gueye e Herrera mais próximos uns dos outros e os avançados

(Messi, Neymar e Mbappé) ligeiramente afastados, possivelmente para transições rápidas. A defesa está mais retraída, o que indica uma abordagem mais cautelosa.

5) Mapa de Passes: Para melhorar a circulação de bola, vais criar um mapa de passes bem-sucedidos e falhados de ambas as equipas. Isto ajudará a identificar lacunas e oportunidades para otimizar a construção de jogo.

Tarefa: Criar um mapa de passes bem-sucedidos e falhados das duas equipas.

Para começar filtrei e organizei os dados dos passes de ambas as equipas:

```
#Organizar dados dos passes por equipa
dados_passes <- dados %>% filter(Action %in% c("Passes accurate", "Passes (inaccurate)"))

dados_passes_city <- dados_passes %>% filter(Team %in% 'Manchester City')

dados_passes_psg <- dados_passes %>% filter(Team %in% 'PSG')
```

Calculei a percentagem de acerto e o número total de passes de cada equipa:

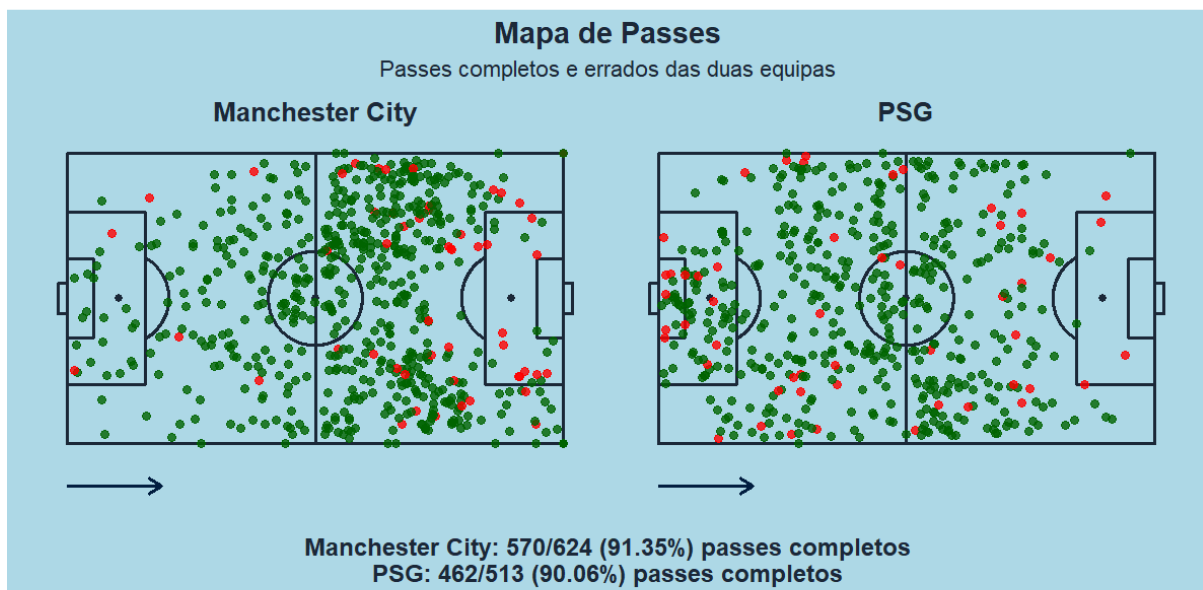
```
# Percentagem geral e número total de passes corretos por equipa
team1_stats <- dados_passes_city %>%
  summarise(
    total_passes = n(),
    accurate_passes = sum(Action == "Passes accurate"),
    accuracy = round((accurate_passes / total_passes) * 100, 2)
  )

team2_stats <- dados_passes_psg %>%
  summarise(
    total_passes = n(),
    accurate_passes = sum(Action == "Passes accurate"),
    accuracy = round((accurate_passes / total_passes) * 100, 2)
  )
```

De seguida elaborei um gráfico para cada equipa que permite verificar onde foram executados os passes e se estes foram bem ou mal sucedidos através de uma legenda:

```
# Gráfico para Manchester City
grafico_team1 <- ggplot() +
  annotate_pitch(
    dimensions = pitch_custom,
    colour = "#1B2838", linewidth = 0.8, fill = "lightblue", limits = FALSE) +
  geom_point(
    data = dados_passes_city,
    aes(x = pos_x, y = pos_y, colour = Action),
    size = 2, alpha = 0.8
  ) +
  scale_color_manual(
    values = c(
      "Passes accurate" = "darkgreen",
      "Passes (inaccurate)" = "red"
    ),
    name = "Tipo de Passe"
  ) +
  annotate(
    "segment",
    x = 0, xend = 20,
    y = -10, yend = -10,
    arrow = arrow(length = unit(0.3, "cm")),
    colour = "#021e40", linewidth = 0.8
  ) +
  labs(
    title = "Manchester City",
    x = NULL,
    y = NULL
  ) +
  theme_pitch() +
```

Combinei os gráficos de ambas as equipes o que resultou no seguinte gráfico:



O mapa de passes revela que ambas as equipes tiveram uma elevada precisão no passe (91.35% para o Manchester City e 90.06% para o PSG), mas o City realizou um número significativamente maior de passes, sugerindo maior domínio da posse de bola.

Os passes falhados concentram-se sobretudo nas alas e na zona ofensiva, o que indica dificuldades na finalização de jogadas e na progressão em espaços mais pressionados.

Para complementar o exercício decidi elaborar um gráfico de barras que demonstra os jogadores com mais passes de cada equipa assim como a percentagem de acerto dos jogadores.

Para tal calculei o top 5 de jogadores com mais passes de cada equipa, assim como a sua percentagem de acerto:

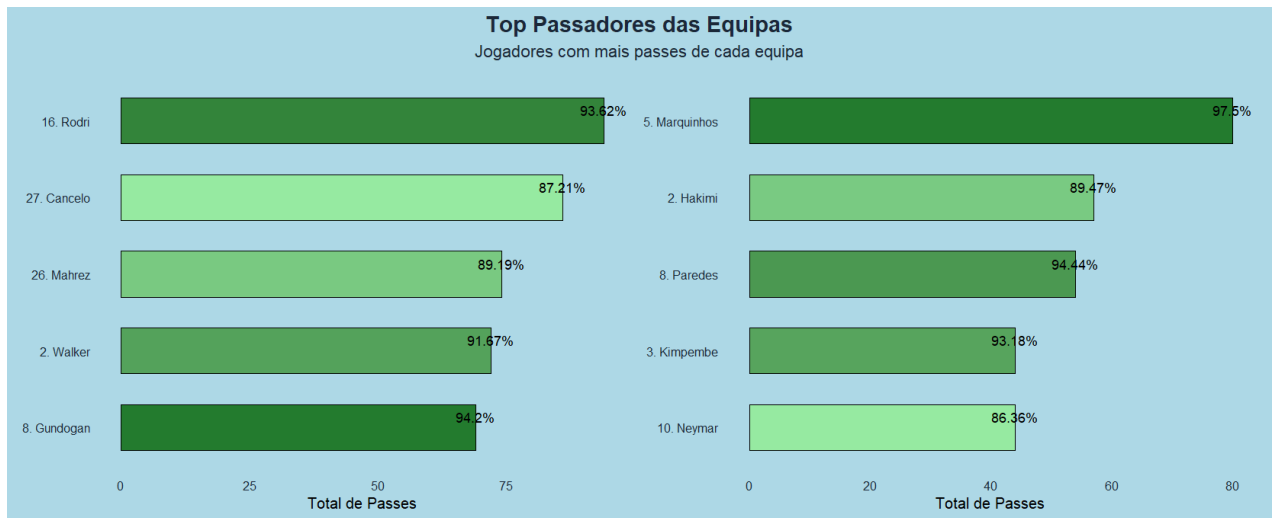
```
# Calcular os top 5 passers de cada equipa
top_passes_city <- dados_passes_city %>%
  group_by(code) %>%
  summarise(total_passes = n(), accurate_passes = sum(Action == "Passes accurate")) %>%
  mutate(accuracy = round((accurate_passes / total_passes) * 100, 2)) %>%
  arrange(desc(total_passes)) %>%
  slice(1:5)

top_passes_psg <- dados_passes_psg %>%
  group_by(code) %>%
  summarise(total_passes = n(), accurate_passes = sum(Action == "Passes accurate")) %>%
  mutate(accuracy = round((accurate_passes / total_passes) * 100, 2)) %>%
  arrange(desc(total_passes)) %>%
  slice(1:5)
```

De seguida elaborei um gráfico de barras para cada uma das equipas com o seu top 5 de jogadores com mais passes em que a cor da barra é definida com base num gradiente que representa a percentagem de passes certos de cada jogador:

```
# Gráfico de barras para top passadores
top_passers_city <- ggplot(top_passes_city, aes(x = reorder(code, total_passes), y = total_passes, fill = accuracy)) +
  geom_bar(stat = "identity", color = "black", alpha = 0.8, width = 0.6) +
  geom_text(aes(label = paste0(accuracy, "%")), vjust = -0.5, size = 3.5, color = "black") +
  scale_fill_gradient(low = "lightgreen", high = "darkgreen", name = "% de Passes completos") +
  labs(
    x = NULL,
    y = "Total de Passes"
  ) +
  coord_flip() +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color = "#1B2838"),
    axis.title.y = element_text(color = "#1B2838"),
    axis.text = element_text(color = "#1B2838"),
    plot.background = element_rect(fill = "lightblue", color = NA),
    panel.background = element_rect(fill = "lightblue", color = NA)
  )
```

Combinei os gráficos de ambas as equipas e o resultado final foi este:



Assim conseguimos não só verificar os jogadores com mais passes de cada equipa, como também a percentagem de passes certos de cada um deles.

6) Mapa de xG: Por fim, vais analisar os golos esperados (xG) das duas equipas. Para impressionares a equipa técnica, calcula a soma total do xG de cada equipa e apresenta esse valor como etiqueta no gráfico. Tarefa: Representar os valores totais de xG das duas equipas.

Para o último exercício, comecei por calcular os xG por minuto com o valor acumulado e também o valor total:

```
# Calcular a soma acumulada de xG para cada equipa por minuto
dados_xg <- dados %>%
  filter(!is.na(xG)) %>%
  mutate(
    start = as.numeric(start),      # Converter start para numérico
    minuto = floor(start / 60)      # Converter para minutos
  ) %>%
  group_by(Team, minuto) %>%        # Agrupar por equipa e minuto
  summarise(
    xG_minuto = sum(xG),            # Somar xG por minuto
    .groups = "drop"
  ) %>%
  arrange(Team, minuto) %>%
  group_by(Team) %>%
  mutate(
    xG_acumulado = cumsum(xG_minuto), # Calcular o acumulado de xG
    xG_total = max(xG_acumulado)      # Calcular o total de xG
  )
```

Depois filtrei as ações que foram golo para ser mais fácil apresentar essa informação no gráfico que pretendia desenvolver:

```

golos <- dados %>%
  filter(Action == "Goals") %>% # Filtrar apenas as ações de golos
  mutate(minuto = floor(as.numeric(start) / 60)) %>% # Arredondar 'start' para minutos
  select(Team, minuto, code)

```

Depois construí o gráfico que apresenta a evolução dos xG ao longo encontro. Adicionei uma imagem com a função `geom_image` nos minutos que foram golo e, além disso apresentei também os jogadores que marcaram os vários golos do jogo. Além disso, adicionei uma etiqueta no final da linha de cada equipa que indica os xG das equipas no final do jogo.

```

grafico_xg <- ggplot(dados_xg, aes(x = minuto, y = xG_acumulado, group = Team, color = Team)) +
  geom_step(size = 1.2) + # Linhas separadas para cada equipa

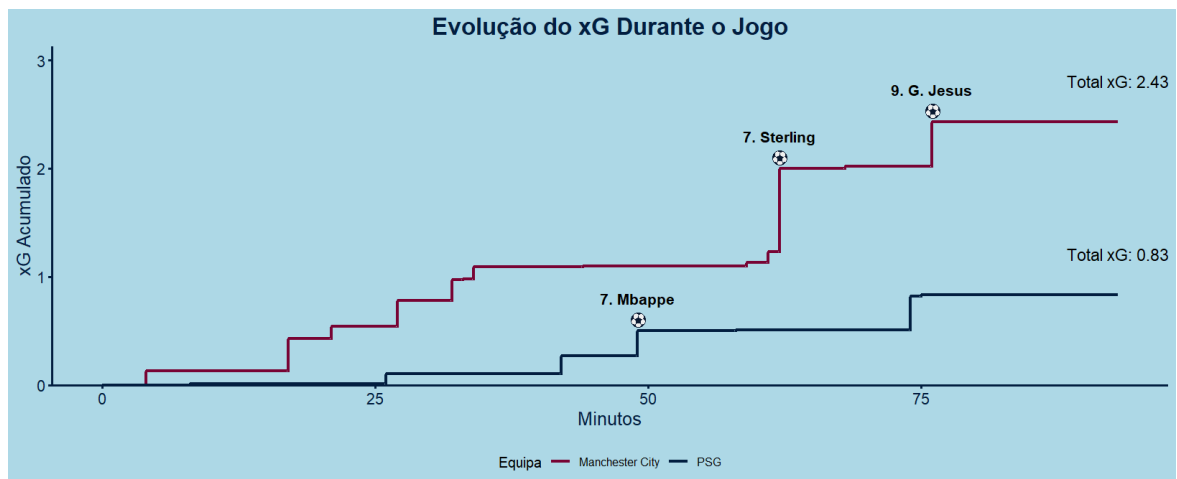
  # Adicionar imagem da bola nos minutos dos golos
  geom_image(
    data = golos %>% left_join(dados_xg, by = c("Team", "minuto")), # Juntar informações de xG acumulado por minuto
    aes(x = minuto, y = xG_acumulado + 0.1, image = "ball.png"), # Adicionar imagem acima da linha
    size = 0.05, inherit.aes = FALSE
  ) +

  # Adicionar etiquetas com o nome do jogador que marcou o golo
  geom_text(
    data = golos %>% left_join(dados_xg, by = c("Team", "minuto")), # Juntar xG acumulado com os golos
    aes(x = minuto, y = xG_acumulado + 0.3, label = code), # Adicionar nome do jogador ao lado da bola
    color = "black", fontface = "bold", size = 4, inherit.aes = FALSE, hjust = 0.5
  ) +

  # Adicionar etiquetas do total de xG ao final das linhas
  geom_text(
    data = dados_xg %>% group_by(Team) %>% summarise(
      x = max(minuto),
      y = max(xG_acumulado),
      label = paste0("Total xG: ", round(max(xG_acumulado), 2))
    ),
    aes(x = x, y = y + 0.2, label = label), # Ajuste vertical com `y + 0.2`
    hjust = 0.5, # Centralizar horizontalmente
    vjust = -1, # Ajustar a posição vertical acima da linha
    size = 4.5, # Reduzir o tamanho do texto
    inherit.aes = FALSE
  ) +

```

O resultado foi o seguinte:



Desenvolvi também outro gráfico. Um mapa que nos permite verificar os xG de cada ação nas várias zonas do terreno.

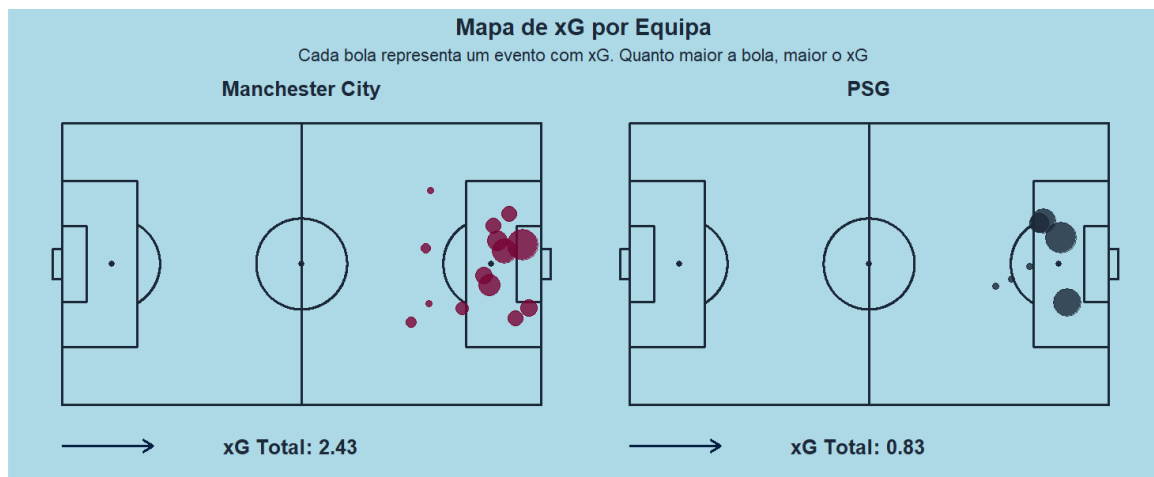
Para isso, filtrei os dados pelas ações com $xG > 0$:

```
# Filtrar apenas os eventos com xG > 0
dados_xg_filtrado <- dados %>%
  filter(xG > 0)
```

E contruí um gráfico em que as ações filtradas são representados com bolas no campo, cujo tamanho varia consoante o valor dos xG:

```
# Criar gráfico para City
grafico_team1 <- ggplot() +
  annotate_pitch(dimensions = pitch_custom, colour = "#182838", linewidth = 0.8, fill = NA, limits = FALSE) +
  geom_point(
    data = dados_xg_filtrado %>% filter(Team == unique(dados_xg_filtrado$Team)[1]),
    aes(x = pos_x, y = pos_y, size = xG,
        color = "#780434", alpha = 0.8)
  ) +
  scale_size_continuous(range = c(2, 10), name = "xG") + # Tamanho das bolas baseado no xG
  labs(
    title = unique(dados_xg_filtrado$Team)[1],
    x = NULL,
    y = NULL
  ) +
  theme_pitch() +
  theme(
    panel.grid = element_blank(),
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, color = "#182838"),
    plot.background = element_rect(fill = "lightblue", color = NA),
    panel.background = element_rect(fill = "lightblue", color = NA),
    legend.position = "none"
  ) +
  annotate(
    "segment",
    x = 0, xend = 20,
    y = -10, yend = -10,
    arrow = arrow(length = unit(0.3, "cm")),
    colour = "#021e40", linewidth = 0.8
  ) +
  annotate(
    "text",
    x = 50, y = -10, # Ajustar para ficar abaixo do gráfico
    label = paste0("xG Total: ", round(xg_totais$xG_total[xg_totais$Team == unique(dados_xg_filtrado$Team)[1]], 2)),
    size = 5, fontface = "bold", color = "#182838"
  )
```

Combinei o gráfico de ambas as equipas e o resultado final foi o seguinte:



Pela análise do gráfico concluímos facilmente que o City dominou o encontro ao apresentar um maior número de oportunidades, maior xG Total e ocasiões mais perigosas que o PSG.