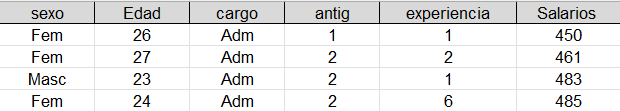
**CONTROL 1 – DIPLOMADO EN MÉTODOS ESTADÍSTICOS**

**NOMBRE: Rodrigo Jeldres Carrasco**

**PARTE 1 (30 ptos)**

La base de datos “Salario1.xlsx” (hoja Salarios) contiene datos de funcionarios de una empresa respecto de las siguientes características:

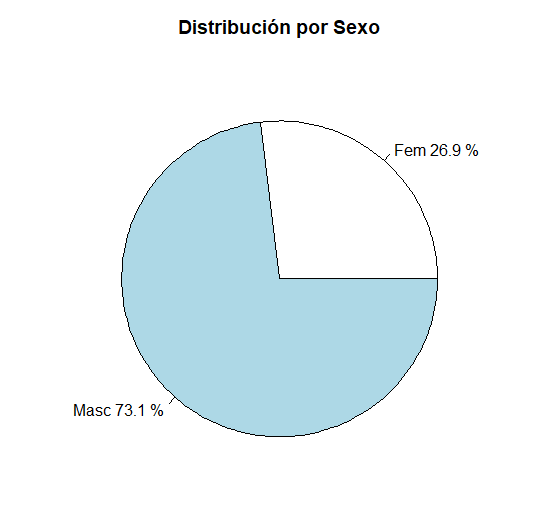
* Sexo de/la funcionaria/o
* Edad (en años)
* Cargos en la empresa (Administrativo, Profesionales, Jefaturas)
* Antigüedad (años dentro de la empresa)
* Experiencia (total de años trabajando en la empresa y otras)
* Salario (ingreso actual)



***Para cada pregunta: copiar/pegar tablas, gráficos y redactar comentarios.***

1. Haga un resumen descriptivo y gráfico de las características de la/os funcionarios:

a.1 Tabla de frecuencia (frecuencias y porcentajes) por sexo y agregue un gráfico circular.

Comentario: visualmente se ve que gran porcentaje de la base es formada por hombres, con un 73,08 para hombres y 26,92% para mujeres

library(readxl)

setwd("d:/dev/Estadistica/")

# Importar la base de datos desde Excel

base <- read\_excel("Control\_1.xlsx", sheet = "Salarios")

attach(base)

tabla\_frecuencia <- table(Sexo)

porcentajes <- prop.table(tabla\_frecuencia) \* 100

tabla\_frecuencia\_df <- data.frame(

Sexo = names(tabla\_frecuencia),

Frecuencia = as.vector(tabla\_frecuencia),

Porcentaje = round(porcentajes, 2)

)

total\_row <- data.frame(

Sexo = "Total",

Frecuencia = total\_frecuencia,

Porcentaje = round(total\_porcentaje, 2)

)

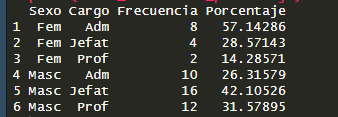
print(tabla\_frecuencia\_df)

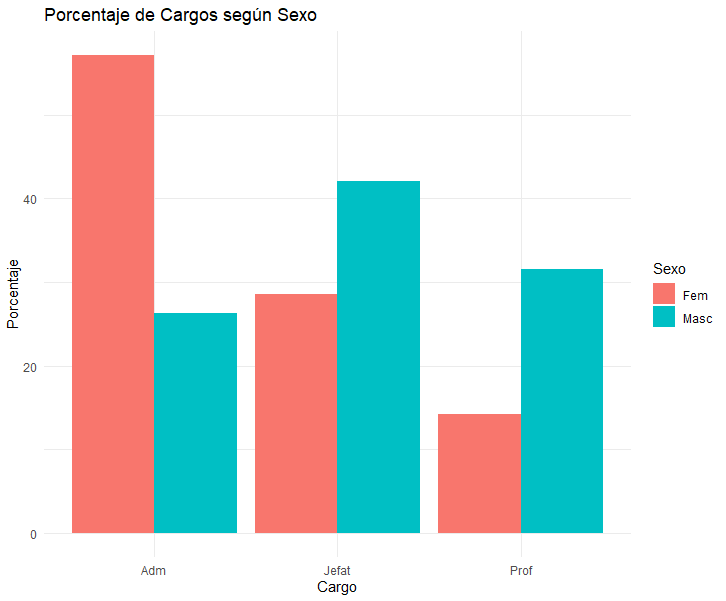
pie(tabla\_frecuencia,

labels = paste(names(tabla\_frecuencia), round(porcentajes, 1), "%"),

main = "Distribución por Sexo")

a.2 Tabla de frecuencia del cargo (frecuencias y porcentajes), según sexo





Comentario: para cargos “administrativos”, se observa mayor presencia femenina, en cambio, para cargos de jefatura, se observa mayor presencia masculina. Misma tendencia de mayor presencia masculina se observa en cargo “Prof” (lo tomaré como profesional)

tabla\_frecuencia\_cargo <- table(Sexo, Cargo)

porcentajes\_cargo <- prop.table(tabla\_frecuencia\_cargo, 1) \* 100

tabla\_frecuencia\_cargo\_df <- as.data.frame(tabla\_frecuencia\_cargo)

colnames(tabla\_frecuencia\_cargo\_df) <- c("Sexo", "Cargo", "Frecuencia")

porcentajes\_cargo\_df <- as.data.frame(porcentajes\_cargo)

colnames(porcentajes\_cargo\_df) <- c("Sexo", "Cargo", "Porcentaje")

tabla\_frecuencia\_porcentaje <- tabla\_frecuencia\_cargo\_df %>%

inner\_join(porcentajes\_cargo\_df, by = c("Sexo", "Cargo"))

tabla\_frecuencia\_porcentaje <- tabla\_frecuencia\_porcentaje %>%

arrange(Sexo, Cargo)

print(tabla\_frecuencia\_porcentaje)

library(ggplot2)

ggplot(tabla\_frecuencia\_porcentaje, aes(x = Cargo, y = Frecuencia, fill = Sexo)) +

geom\_bar(stat = "identity", position = "dodge") +

labs(title = "Frecuencia de Cargos según Sexo", x = "Cargo", y = "Frecuencia") +

theme\_minimal()

ggplot(tabla\_frecuencia\_porcentaje, aes(x = Cargo, y = Porcentaje, fill = Sexo)) +

geom\_bar(stat = "identity", position = "dodge") +

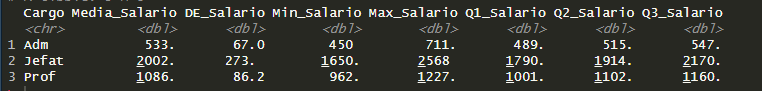
labs(title = "Porcentaje de Cargos según Sexo", x = "Cargo", y = "Porcentaje") +

theme\_minimal()

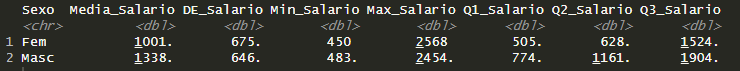
a.3 Resumen numérico de edad, salario por sexo y salario por cargo

Redacte un resumen/comentario dada la información obtenida en los puntos anteriores.

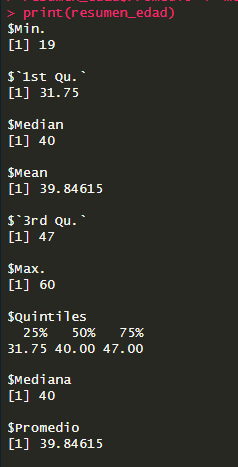
Para salario por cargo:



Para salario por sexo



Resumen numérico de la edad



Comentarios:

* Diferencias Salariales por Cargo: se ve una clara jerarquía de sueldo entre los cargos, con los jefes ganando significativamente más que los profesionales y administrativos. La variabilidad de los salarios es mayor en los cargos de jefatura, (supongo que podría ser por experiencia o tiempo en la empresa lo que empuja el sueldo al alza)
* Diferencias Salariales por Sexo: Los hombres tienen un salario promedio superior al de las mujeres. La distribución salarial de las mujeres muestra una mayor dispersión, lo que podría indicar una mayor presencia de mujeres en posiciones con salarios más bajos. La mediana salarial de las mujeres es más baja que la de los hombres, lo que podría indicar posibles brechas salariales entre géneros.
* Desviación Estándar: La variabilidad de los salarios también es alta, tanto para el caso de agrupar por sexo como por cargo
* Los hombres tienen un salario promedio superior al de las mujeres

resumen\_edad <- summary(Edad)

resumen\_edad$Quintiles <- quantile(Edad, probs = c(0.25, 0.5, 0.75))

resumen\_edad$Mediana <- median(Edad)

resumen\_edad$Promedio <- mean(Edad)

print(resumen\_edad)

datos\_agrupados\_cargo <- base %>%

group\_by(Cargo) %>%

summarize(Media\_Salario = mean(Salarios),

DE\_Salario = sd(Salarios),

Min\_Salario = min(Salarios),

Max\_Salario = max(Salarios),

Q1\_Salario = quantile(Salarios, probs = 0.25),

Q2\_Salario = median(Salarios),

Q3\_Salario = quantile(Salarios, probs = 0.75))

print(datos\_agrupados\_cargo)

datos\_agrupados\_sexo <- base %>%

group\_by(Sexo) %>%

summarize(Media\_Salario = mean(Salarios),

DE\_Salario = sd(Salarios),

Min\_Salario = min(Salarios),

Max\_Salario = max(Salarios),

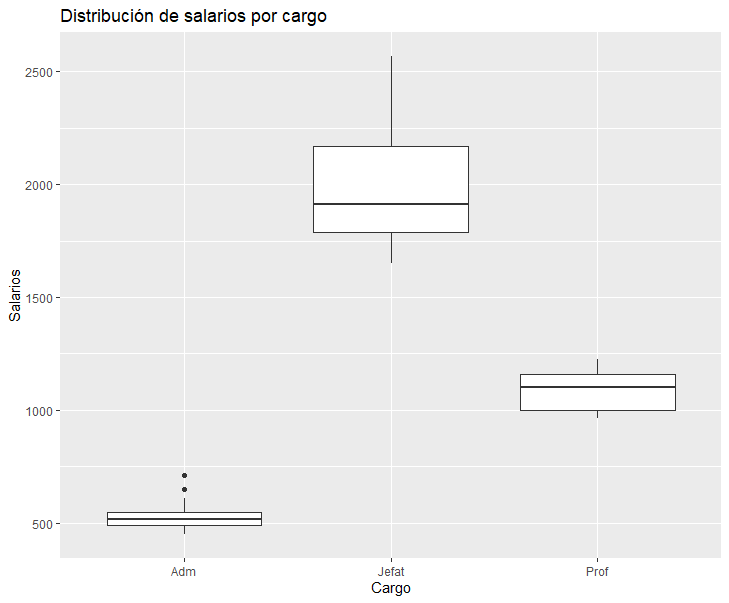
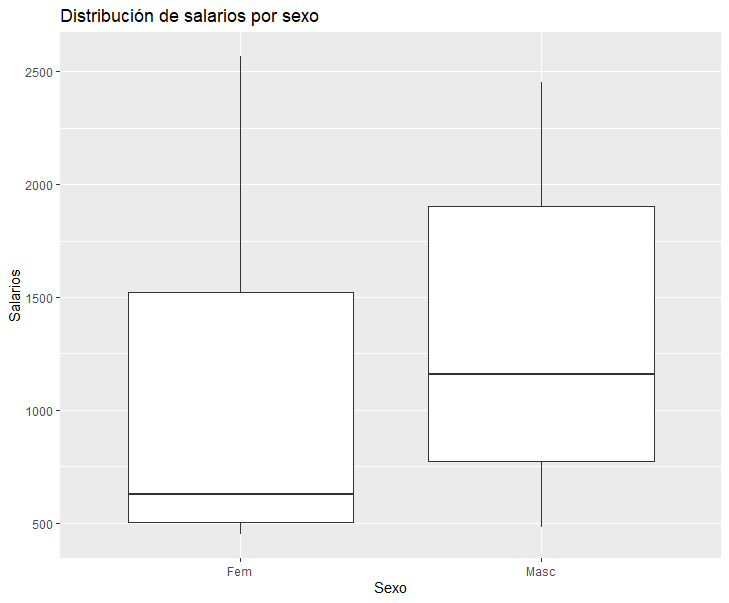
Q1\_Salario = quantile(Salarios, probs = 0.25),

Q2\_Salario = median(Salarios),

Q3\_Salario = quantile(Salarios, probs = 0.75))

print(datos\_agrupados\_sexo)

1. Haga gráficos de caja de la distribución de los salarios según el cargo y según el sexo de la/os funcionarios. Escriba comentarios sobre los salarios para ambos casos.

Comentario

* Para el caso de grafico salario/sexo, la mediana de género “masculino” es mayor a la mediana de género “femenino”
* Para el caso administrativo, se tienen dos casos atípicos
* Con el grafico por cargos se observa una marcada tendencia al alza en cargos de jerarquía , es decir, administrativo está por debajo de profesional, y profesional estaá por debajo de “Jefatura”

ggplot(base, aes(x = Cargo, y = Salarios)) +

geom\_boxplot() +

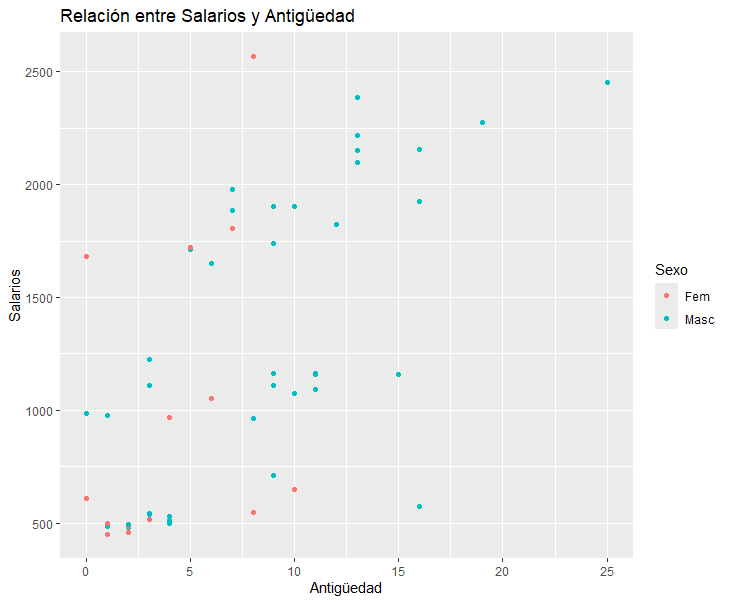
labs(title = "Distribución de salarios por cargo", x = "Cargo", y = "Salarios")

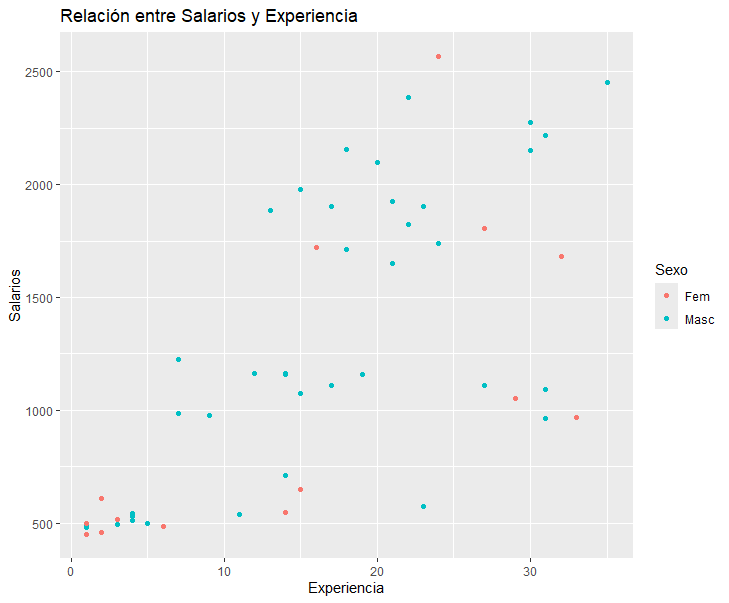
ggplot(base, aes(x = Sexo, y = Salarios)) +

geom\_boxplot() +

labs(title = "Distribución de salarios por sexo", x = "Sexo", y = "Salarios")

1. ¿Hay alguna relación entre Salarios y Antigüedad? ¿Hay alguna relación entre Salarios y Experiencia? Haga gráficos de dispersión (incluya separación por sexo dentro de cada gráfico). Redacte algún comentario.





Comentario:

Se observa ausencia de relación lineal. Los puntos se distribuyen de manera aleatoria por todo el gráfico, esto sugiere que no existe una relación lineal significativa entre las variables. No hay una tendencia clara a que el salario aumente o disminuya a medida que aumenta la antigüedad o la experiencia.

ggplot(base, aes(x = Antig, y = Salarios, color = Sexo)) +

geom\_point() +

labs(title = "Relación entre Salarios y Antigüedad", x = "Antigüedad", y = "Salarios")

ggplot(base, aes(x = Experiencia, y = Salarios, color = Sexo)) +

geom\_point() +

labs(title = "Relación entre Salarios y Experiencia", x = "Experiencia", y = "Salarios")

**CONTROL 1 – DIPLOMADO EN MÉTODOS ESTADÍSTICOS**

**NOMBRE: Rodrigo Jeldres Carrasco**

**PARTE 2 (30 ptos)**

La base de datos “Salario1.xlsx” (hoja Espectaculos) contiene datos de diferentes tipos de espectáculos (deportivos, teatro...) realizados en diferentes regiones del país que fueron realizados en varios establecimientos durante un año determinado (los datos corresponden al año 2016). Importar la base en R.

***Para cada pregunta: copiar/pegar tablas, gráficos y redactar comentarios.***

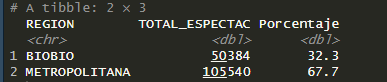
1. Elija dos regionesde la base de datos con las que va a trabajar y guarde en una nueva base.

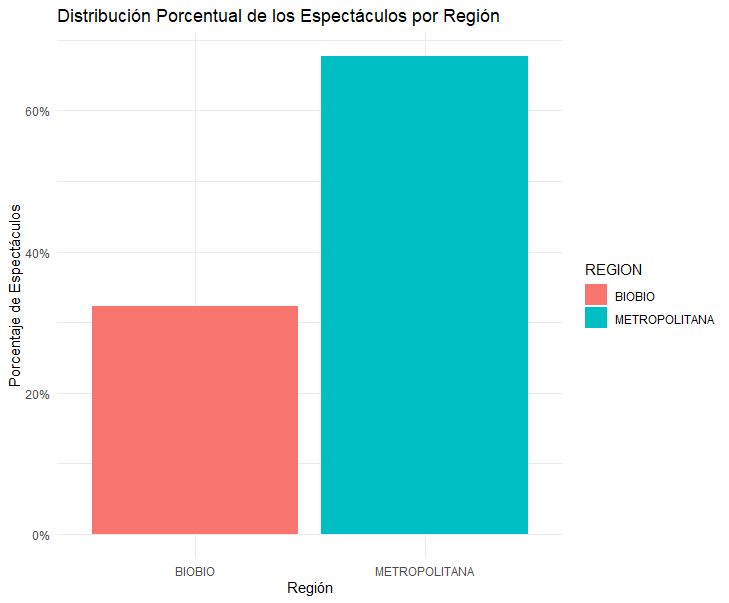
Ayuda: Puede utilizar el comando filter de la siguiente manera:

library(dplyr)

basefinal <-filter(baseinicial,nombrecolumna==" " | " ")

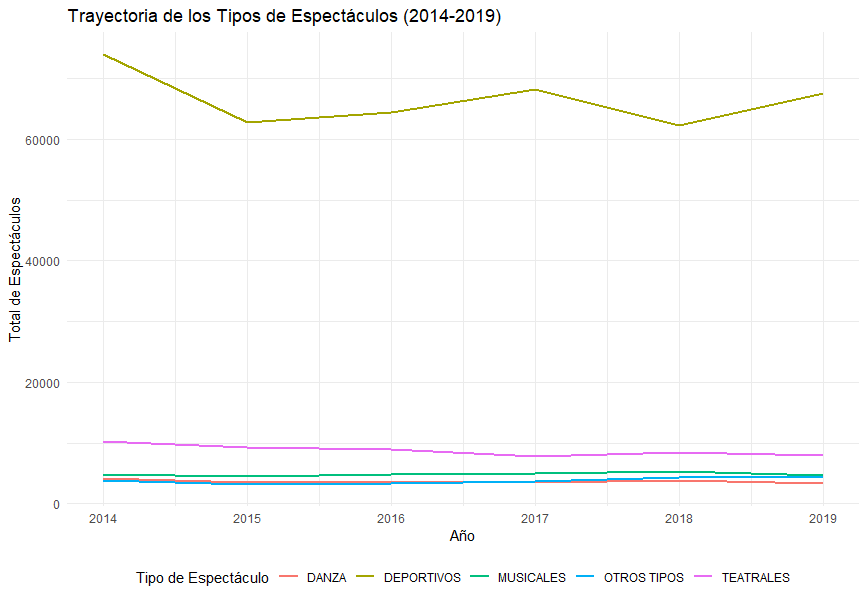
Muestre en una tabla y un gráfico de barrasque muestre la distribución porcentual de los espectáculos, según cada región. Comente la distribución.





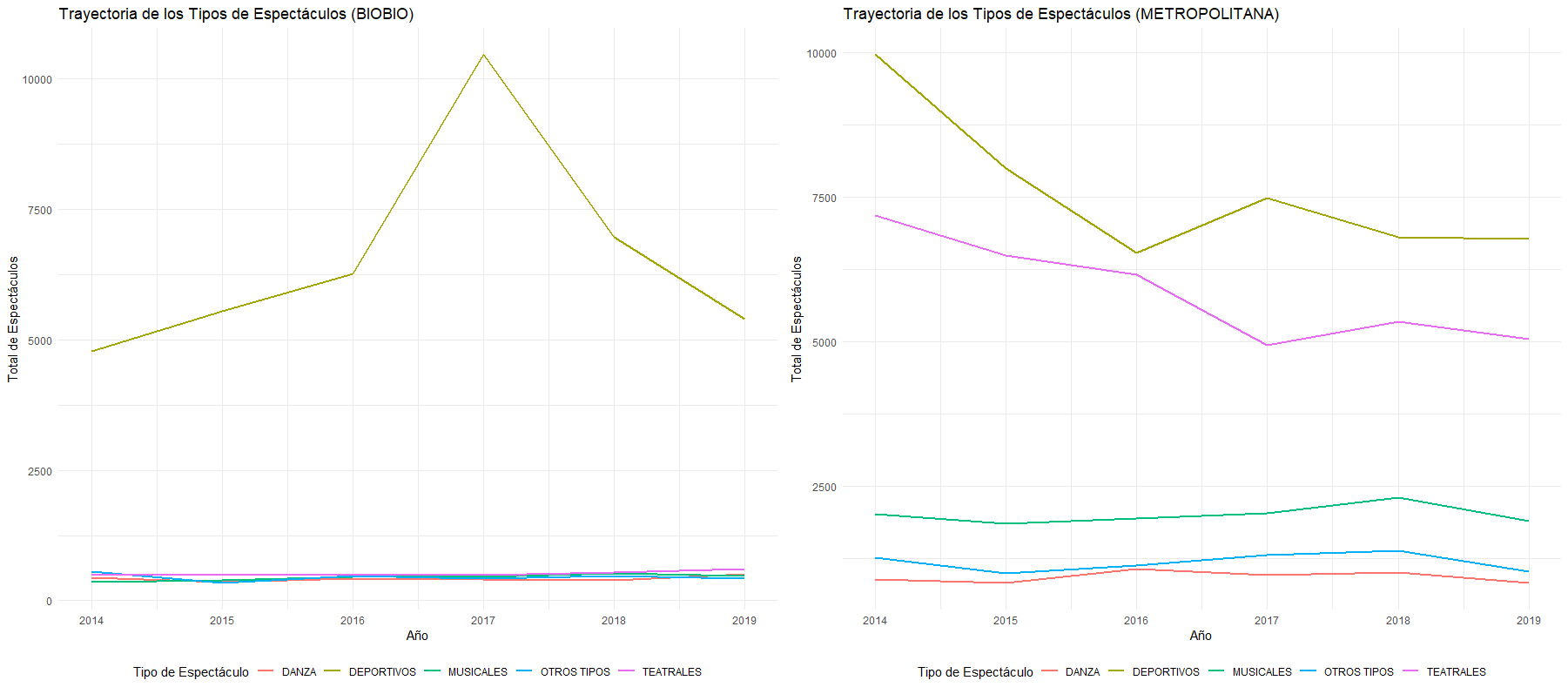
* Comentario: La Región Metropolitana tiene mayor escena de espectáculos con un 67.7% del total, mientras que la Región Biobío solo tiene el 32.3%. Esta disparidad podría ser por una centralización de actividades en la capital
* La mayor cantidad de espectáculos en la Región Metropolitana, podría explicarse a diversos factores como la mayor población o infraestructura en comparación con la Región Biobío.

1. Muestre la trayectoria, con un gráfico de líneas, de los tipos de espectáculos en el transcurso de los años (2014-2019) sin separar por regiones. Escriba comentarios.



Comentario: Se observa que, con mucha diferencia al alza, los **espectáculos deportivos** son los más realizados en todas las regiones, y los más realizados respecto a otros espectáculos

1. Muestre la trayectoria, con un gráfico de líneas, de los tipos de espectáculos en el transcurso de los años (2014-2019) para cada una de las dos regiones que seleccionó en la parte a). Coloque los dos gráficos en una solo imagen. Escriba comentarios. Ayuda: Utilice grid.arrange(graf1, graf2, ncol = 2) para combinar dos gráficos ggplot en un mismo cuadro



Comentario: para ambas regiones (metropolitana y Biobío) se observa como tendencia a tener mayor cantidad de espectáculos deportivos, esto se mantiene durante el rango de los años 2014 a 2019

Comentario: para ambas regiones, la menor cantidad de espectáculos corresponde a la categoría “danza”