



# Describiendo datos

# Recordemos las columnas

## ■ Una tabla de datos

- hoy en día “planilla”, “matriz”, **data.frame** o tibble en R, entre otros

folio: identificador del hogar

o: número de orden de la persona dentro del hogar

id.vivienda: identificador de la vivienda (comp: comuna área seg viv)

hogar: identificación del hogar en la vivienda

region: región

provincia: provincia

comuna: comuna

ing.comuna: posición en el ranking histórico del ingreso de la comuna (ascendente)

zona: área geográfica (Urbano, Rural)

sexo: sexo de la persona registrada

edad: edad de la persona registrada

ecivil: estado civil de la persona registrada

ch1: situación ocupacional de la persona registrada

ytot: ingreso total

# Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?

- primero debemos **tabular** los datos

```
dfm <- as.data.frame(table(muestra[["provincia"]]))  
colnames(dfm) <- c("Provincia", "Frecuencia")
```

- veamos lo que conseguimos:

```
> dfm  
  Provincia Frecuencia  
1 Chacabuco      5  
2 Cordillera      5  
3 Maipo          7  
4 Melipilla       3  
5 Santiago        76  
6 Talagante       4  
> str(dfm)  
'data.frame':   6 obs. of  2 variables:  
 $ Provincia : Factor w/ 6 levels "Chacabuco", "Cordillera", ... : 1 2 3 4 5 6  
 $ Frecuencia: int  5 5 7 3 76 4
```

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - sirve un **gráfico de barras**

```
p1 <- ggbarplot(  
  dfm,  
  x = "Provincia", y = "Frecuencia",  
  label = TRUE, lab.pos = "out", lab.col = "black",  
  fill = "Provincia", palette = "jco",  
  title = "Muestra de entrevistados en la\nRegión Metropolitana por provincia"  
)  
p1 <- p1 + rotate_x_text(angle = 45)  
  
> p1
```



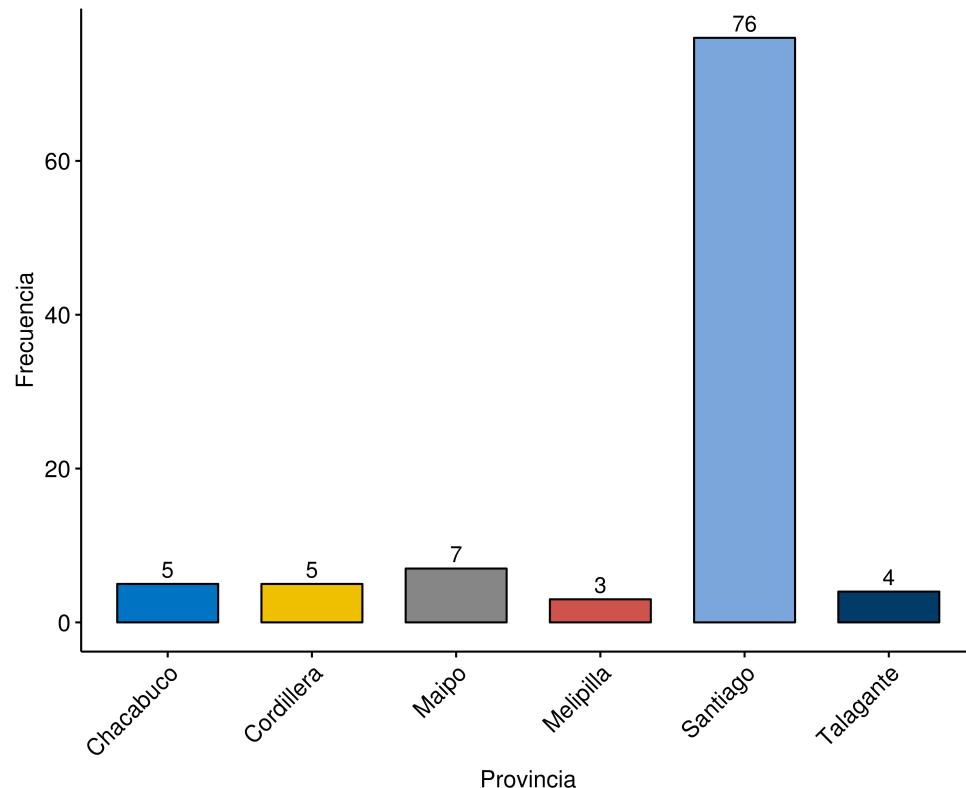
## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - sirve un **gráfico de barras**

Muestra de entrevistados en la Región Metropolitana por provincia

Provincia

Chacabuco	Maipo	Santiago
Cordillera	Melipilla	Talagante



## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - también sirve un **gráfico de torta**

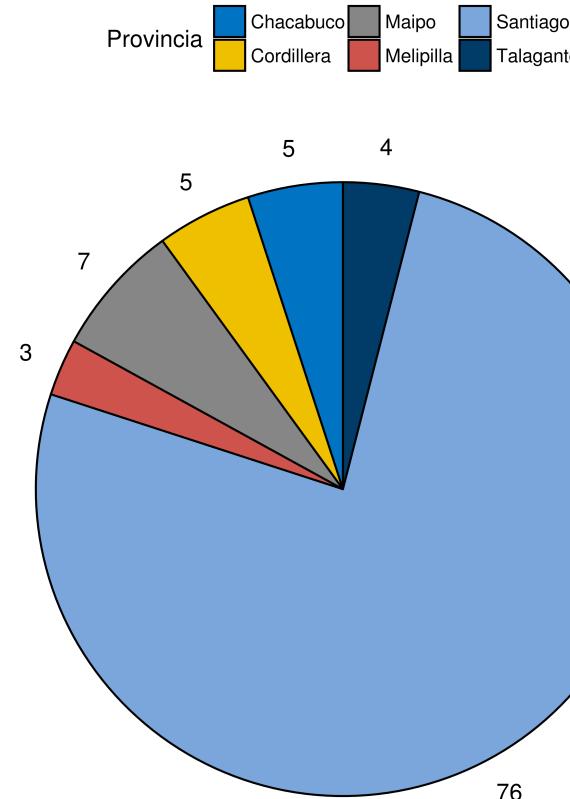
```
p2 <- ggpie(  
  dfm,  
  x = "Frecuencia", label = "Frecuencia",  
  lab.pos = "out", lab.col = "black",  
  fill = "Provincia", color = "black", palette = "jco",  
  title = "Muestra de entrevistados en la\nRegión Metropolitana por  
provincia"  
)  
  
> p2
```



## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - también sirve un **gráfico de torta**

Muestra de entrevistados en la Región Metropolitana por provincia



## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - los gráficos claramente muestran que no
  - ¿cómo representar esto en una medida?

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - los gráficos claramente muestran que no
  - ¿cómo representar esto en una medida?
  - de hecho, la **frecuencia** es una medida que usa para estos efectos
  - aunque la medida más estudiada es **la proporción**, puesto que es un **estimador** de la **probabilidad**

## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - calculemos las proporciones de encuestados en la provincia de Santiago versus otras provincias

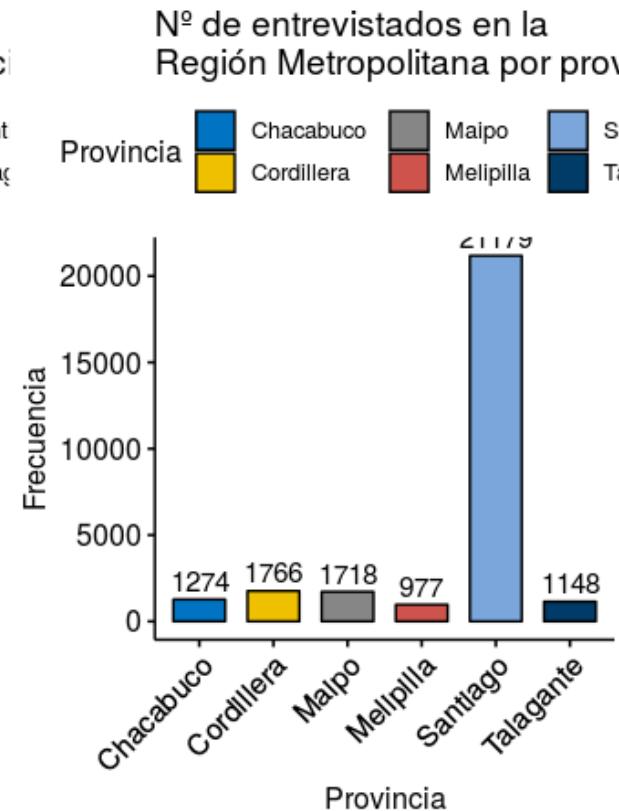
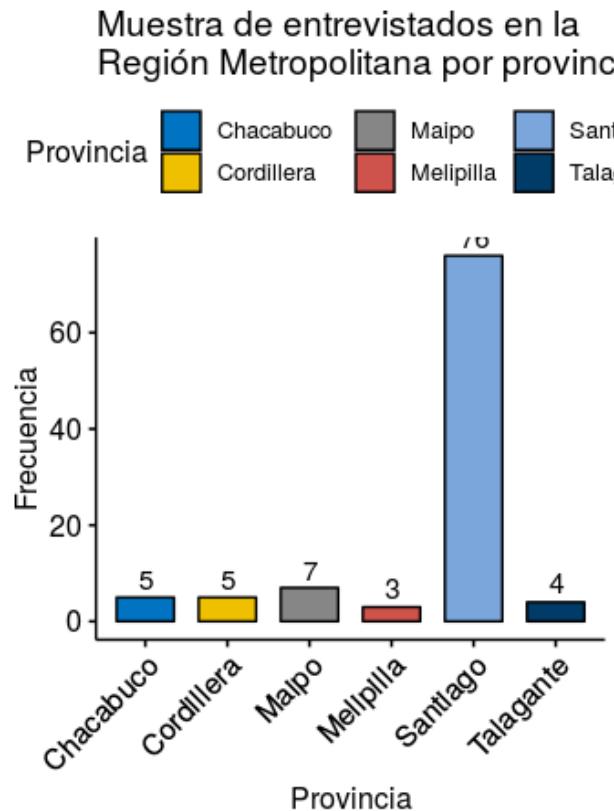
```
Proporciones.muestra <- dfm[["Frecuencia"]] / tamaño.muestra
i <- dfm[["Provincia"]] == "Santiago"
dfm.prop <- data.frame(
  Provincia = c("Santiago", "Otra provincia"),
  Proporción = c(Proporciones.muestra[i], sum(Proporciones.muestra[!i])))
)

> dfm.prop
  Provincia Proporción
1  Santiago      0.76
2 Otra provincia     0.24
```



## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - en la **población**, ¿se observa un comportamiento similar al observado en la muestra?



## Analizando frecuencias

- ¿Se encuestaron más o menos la misma **cantidad de gente** en cada provincia de la RM?
  - en la **población**, ¿se observa un comportamiento similar al observado en la muestra?

Proporciones en la muestra ( $n = 100$ )

---

	Provincia	Proporción
1	Santiago	0.76
2	Otra provincia	0.24

Proporciones en la población ( $n = 28062$ )

---

	Provincia	Proporción
1	Santiago	0.75
2	Otra provincia	0.25

# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - en la **población**, ¿se observa un comportamiento similar al observado en la muestra?
  - vemos que **sí** se representa bien
    - tal vez con la excepción de la provincia Cordillera que se ve un tanto subrepresentada en la muestra

# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - en este caso es una **variable continua** (o casi)
  - una buena alternativa para este caso es un **histograma**
  - pero antes, vamos a cambiar la escala para tener números más claros

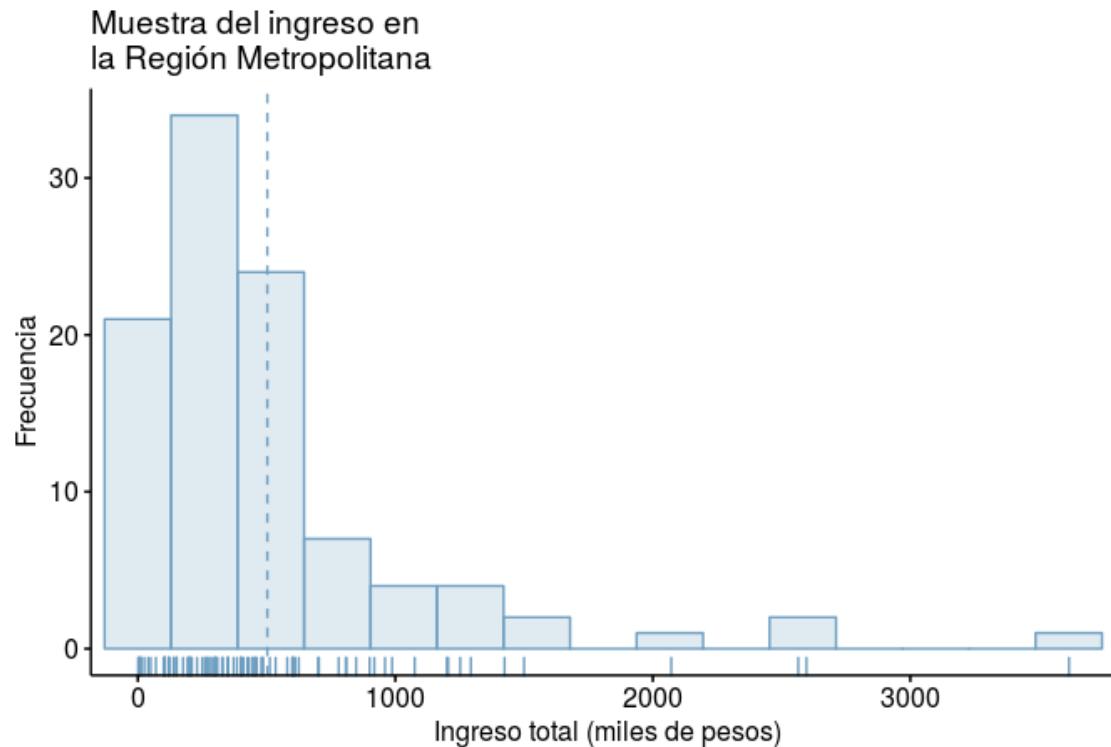
```
m <- muestra
m[["ytot"]] <- m[["ytot"]] / 1000    # cambia escala de $1 a $1.000

p5 <- gghistogram(
  m, x = "ytot", bins = 15,
  add = "mean", rug = TRUE,
  color = "#6D9EC1", fill = "#BFD5E3",
  title = "Muestra del ingreso en\nla Región Metropolitana",
  xlab = "Ingreso total (miles de pesos)", ylab = "Frecuencia"
)
p5 <- ggpar(p5, xlim = c(0, 3650))
```



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - en este caso es una **variable continua** (o casi)
  - una buena alternativa para este caso es un **histograma**



# Analizando una variable numérica

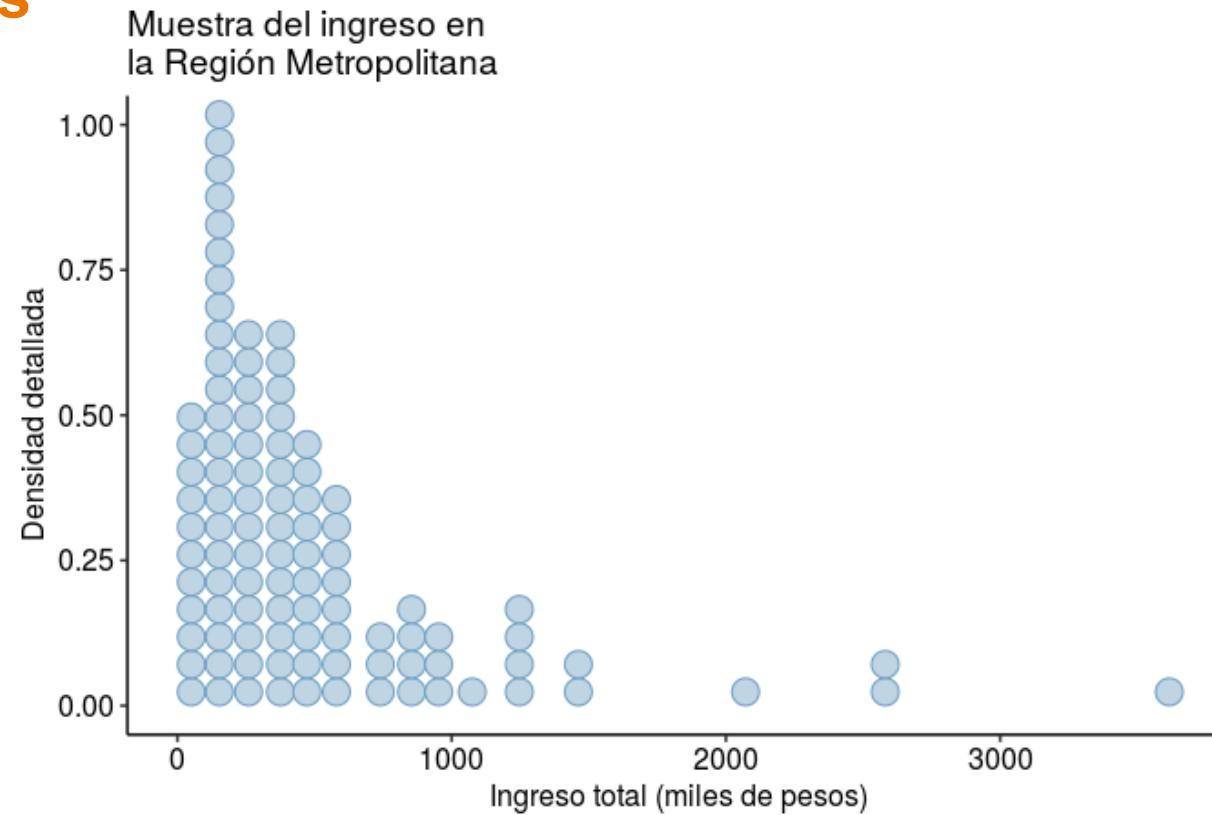
- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - hay alternativas más detallada, como un **diagrama de puntos** o un polígono (o área) de frecuencias
  - pero debemos usar funciones para graficar más básicas, que implementa el **paquete ggplot2**

```
p6 <- ggplot(m, aes(x = ytot))
p6 <- p6 + geom_dotplot(
  binwidth = 100,
  color = "#6D9EC1", fill = "#BFD5E3",
)
p6 <- p6 + xlim(0, 3650)
p6 <- p6 + ggtitle("Muestra del ingreso en\nla Región Metropolitana")
p6 <- p6 + xlab("Ingreso total (miles de pesos)")
p6 <- p6 + ylab("Densidad detallada")
p6 <- p6 + theme_pubr()
```



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - hay alternativas más detallada, como un **diagrama de puntos**



# Analizando una variable numérica

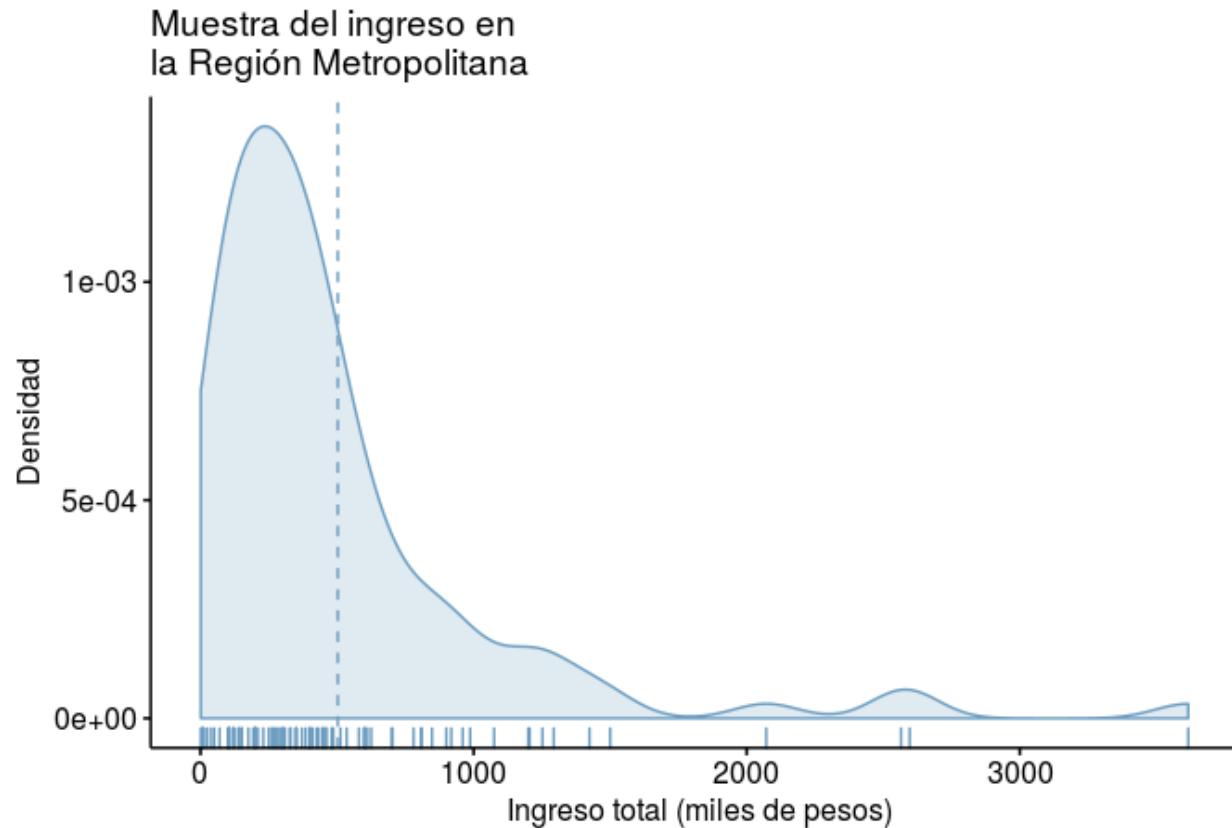
- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - otra muy buena alternativa es **diagrama de densidad**

```
p7 <- ggdensity(  
  m,  
  x = "ytot",  
  add = "mean", rug = TRUE,  
  color = "#6D9EC1", fill = "#BFD5E3",  
  title = "Muestra del ingreso en\nla Región Metropolitana",  
  xlab = "Ingreso total (miles de pesos)", ylab = "Densidad"  
)  
p7 <- ggpar(p7, xlim = c(0, 3650))
```



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - otra muy buena alternativa es **diagrama de densidad**



# Analizando una variable numérica

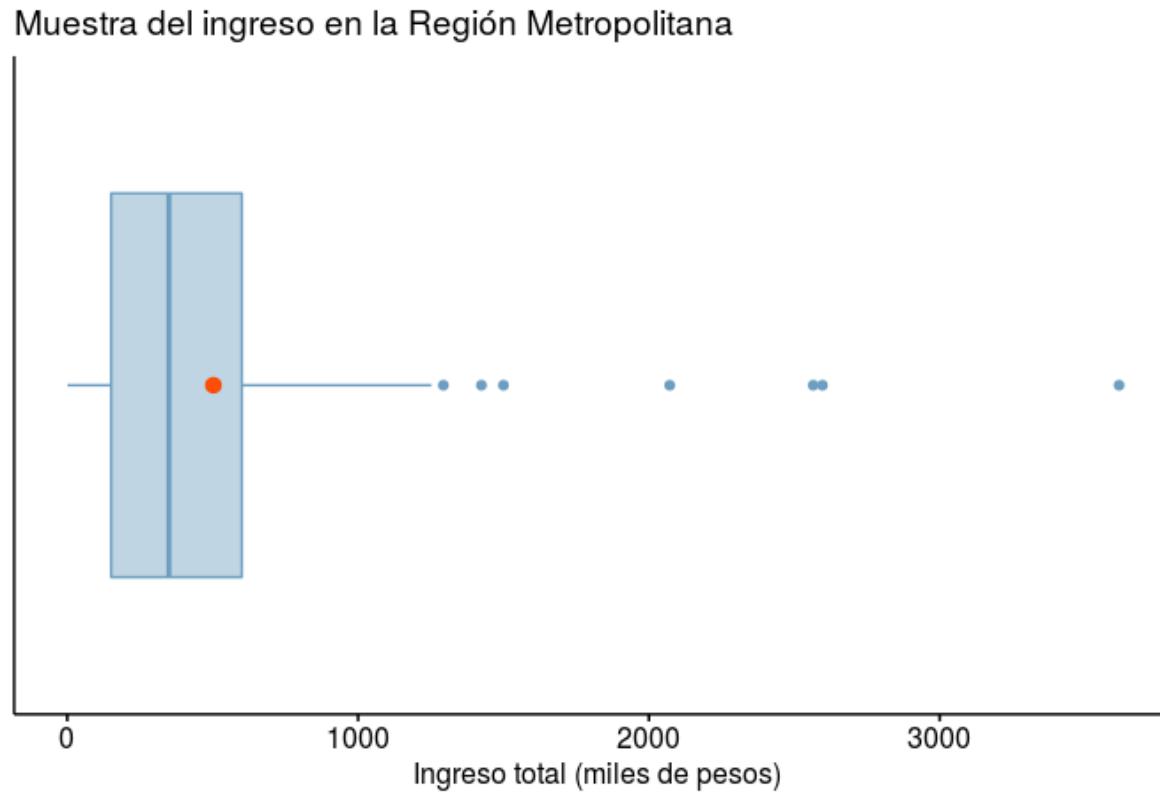
- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - o esconder más detalles usando un **diagrama de cajas**
    - al cual le agregamos un punto con la media
    - (necesitamos crear y ocultar un único grupo)

```
m[["grp"]] <- 1

p8 <- ggboxplot(
  m, x = "grp", y = "ytot",
  add = "mean", add.params = list(color = "#FC4E07"),
  color = "#6D9EC1", fill = "#BFD5E3",
  title = "Muestra del ingreso en la Región Metropolitana",
  ylab = "Ingreso total (miles de pesos)"
)
p8 <- p8 + scale_y_continuous(limits = c(0, 3650))
p8 <- ggrepel(p8, ylab = FALSE, orientation = "horizontal")
p8 <- p8 + rremove("y.text")
p8 <- p8 + rremove("y.ticks")
```

# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - o esconder más detalles usando un **diagrama de cajas**
    - al cual le agregamos un punto con la media



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - o esconder más detalles usando un **diagrama de cajas**
    - o arrepentirnos y ver los datos que generan la caja

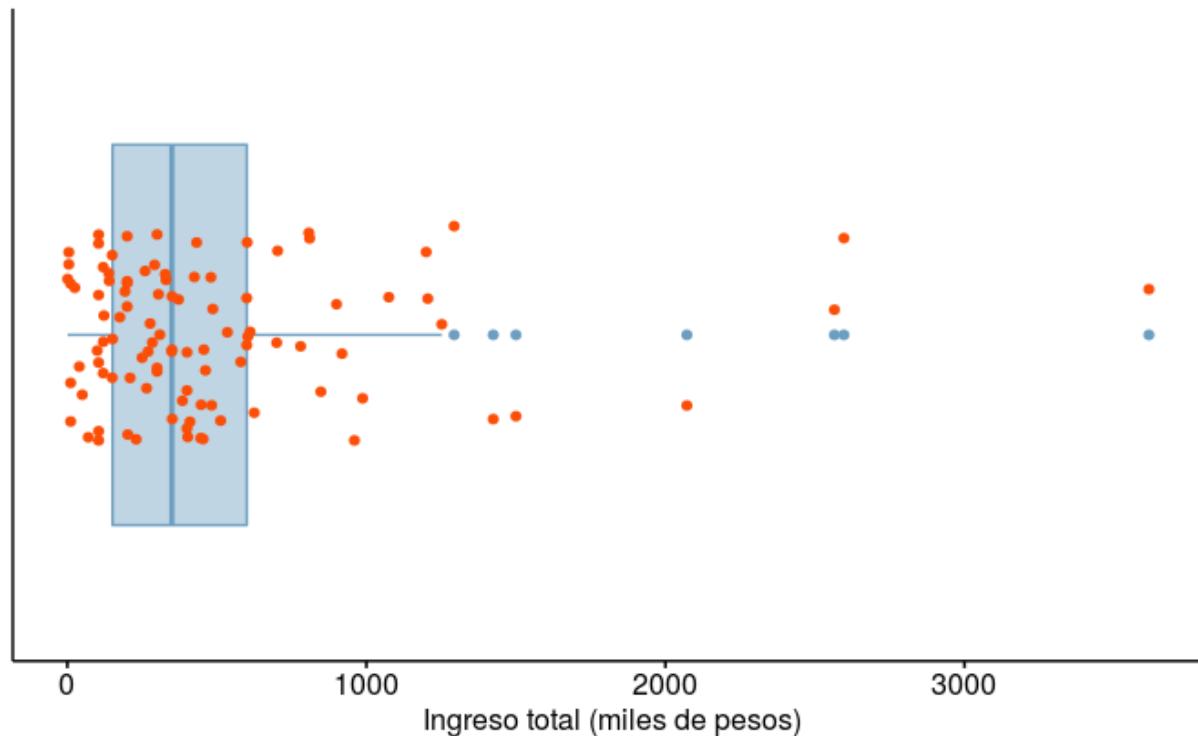
```
p9 <- ggboxplot(  
  m, x = "grp", y = "ytot",  
  add = "jitter", add.params = list(color = "#FC4E07"),  
  color = "#6D9EC1", fill = "#BFD5E3",  
  title = "Muestra del ingreso en la Región Metropolitana",  
  ylab = "Ingreso total (miles de pesos)"  
)  
p9 <- p9 + scale_y_continuous(limits = c(0, 3650))  
p9 <- ggpar(p9, ylab = FALSE, orientation = "horizontal")  
p9 <- p9 + rremove("y.text")  
p9 <- p9 + rremove("y.ticks")
```



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - o esconder más detalles usando un **diagrama de cajas**
  - o arrepentirnos y ver los datos que generan la caja

Muestra del ingreso en la Región Metropolitana



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - el ingreso muestra una **distribución asimétrica**
    - existen valores **atípicamente** altos (outliers)
    - que **distorsionan la representatividad** de medidas como la media

```
> mean(muestra[["ytot"]])
[1] 502712.5
```
    - luego deberíamos usar medidas **más robustas**, tales como:
      - mediana

```
> median(muestra[["ytot"]])
[1] 349500
```
      - rango intercuartil (IQR)

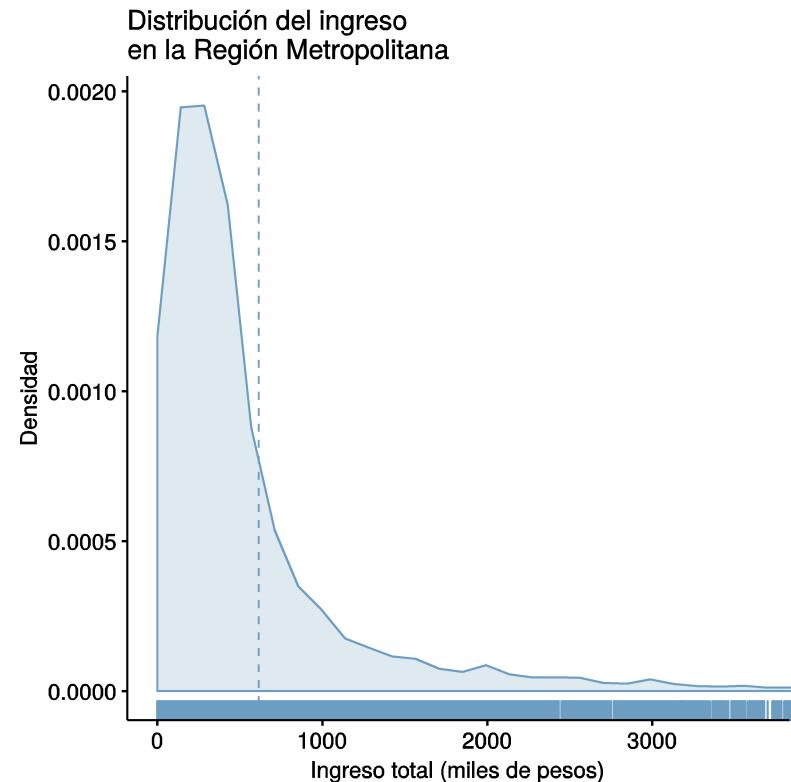
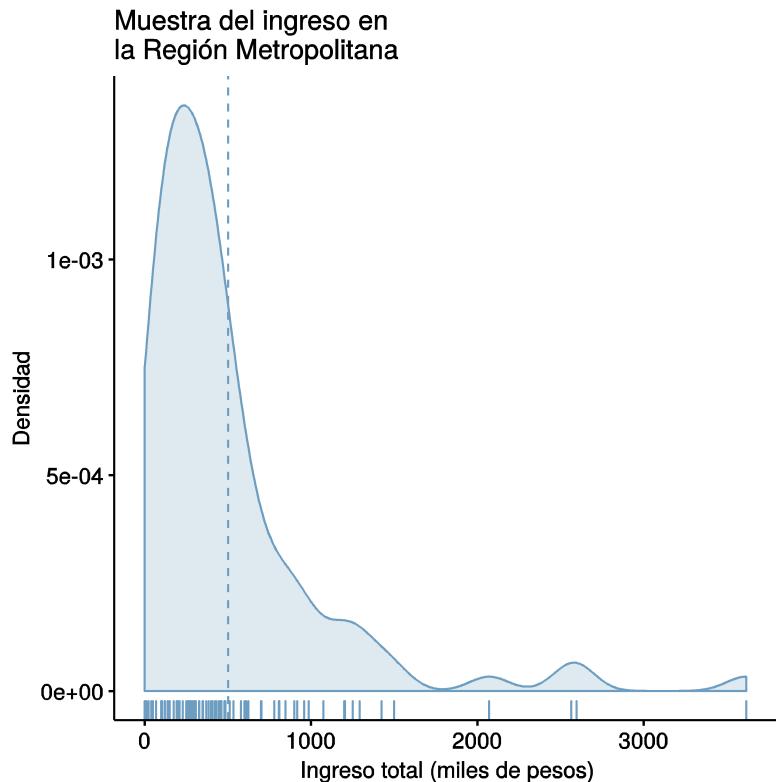
```
> IQR(muestra[["ytot"]])
[1] 449375
```
      - media truncada (trimmed), aquí al 60% de los datos

```
> mean(muestra[["ytot"]], trim = 0.2)
[1] 357332
```



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - en la **población**, ¿se observa un comportamiento similar al observado en la muestra?



# Analizando una variable numérica

- ¿Cómo podemos **describir el ingreso** de los chilenos de la RM?
  - en la **población**, ¿se observa un comportamiento similar al observado en la muestra?
    - como supusimos, las medias tienen una gran diferencia:
      - Media de la muestra: 502712.5
      - Media de la población: 614722.4
    - mientas que las medidas más robustas son más similares:
      - Mediana de la muestra: 349500
      - Mediana de la población: 350000
      - Rango intercuartil de la muestra: 449375
      - Rango intercuartil de la población: 456000
      - Media trucada (20%) de la muestra: 357332
      - Media trucada (20%) de la población: 363930.6

# Analizando una variable numérica en grupos distintos

- ¿Tienen hombres y mujeres **ingresos similares**?
  - nos sirven los mismos gráficos de la pregunta anterior
    - pero mostrando cada grupo **por separado**, y así
    - poder comparar la “**masa**” o/y la “**forma**” de los datos
    - basta usar la **variable de agrupación** como **color** del gráfico
    - en jerga de estadísticos, esta variable categórica se llama **factor**

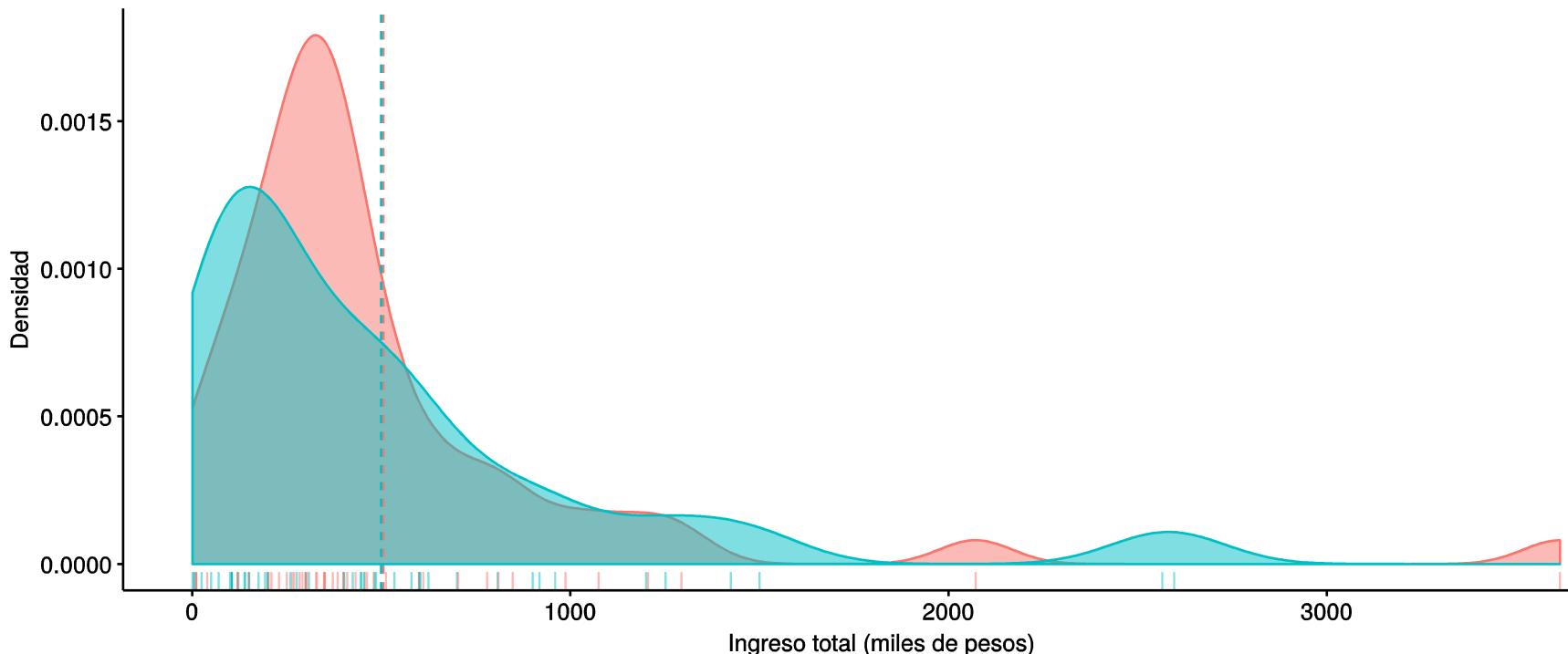
```
p12 <- ggdensity(  
  m, x = "ytot",  
  add = "mean", rug = TRUE,  
  color = "sexo", fill = "sexo",  
  title = "Muestra del ingreso en la Región Metropolitana",  
  subtitle = "(distribuciones separadas por sexo)",  
  xlab = "Ingreso total (miles de pesos)", ylab = "Frecuencia"  
)  
p12 <- ggpar(p12, xlim = c(0, 3650))
```

# Analizando una variable numérica en grupos distintos

- ¿Tienen hombres y mujeres **ingresos similares**?
  - nos sirven los mismos gráficos de la pregunta anterior

Muestra del ingreso en la Región Metropolitana  
(distribuciones separadas por sexo)

sexo ■ Hombre ■ Mujer



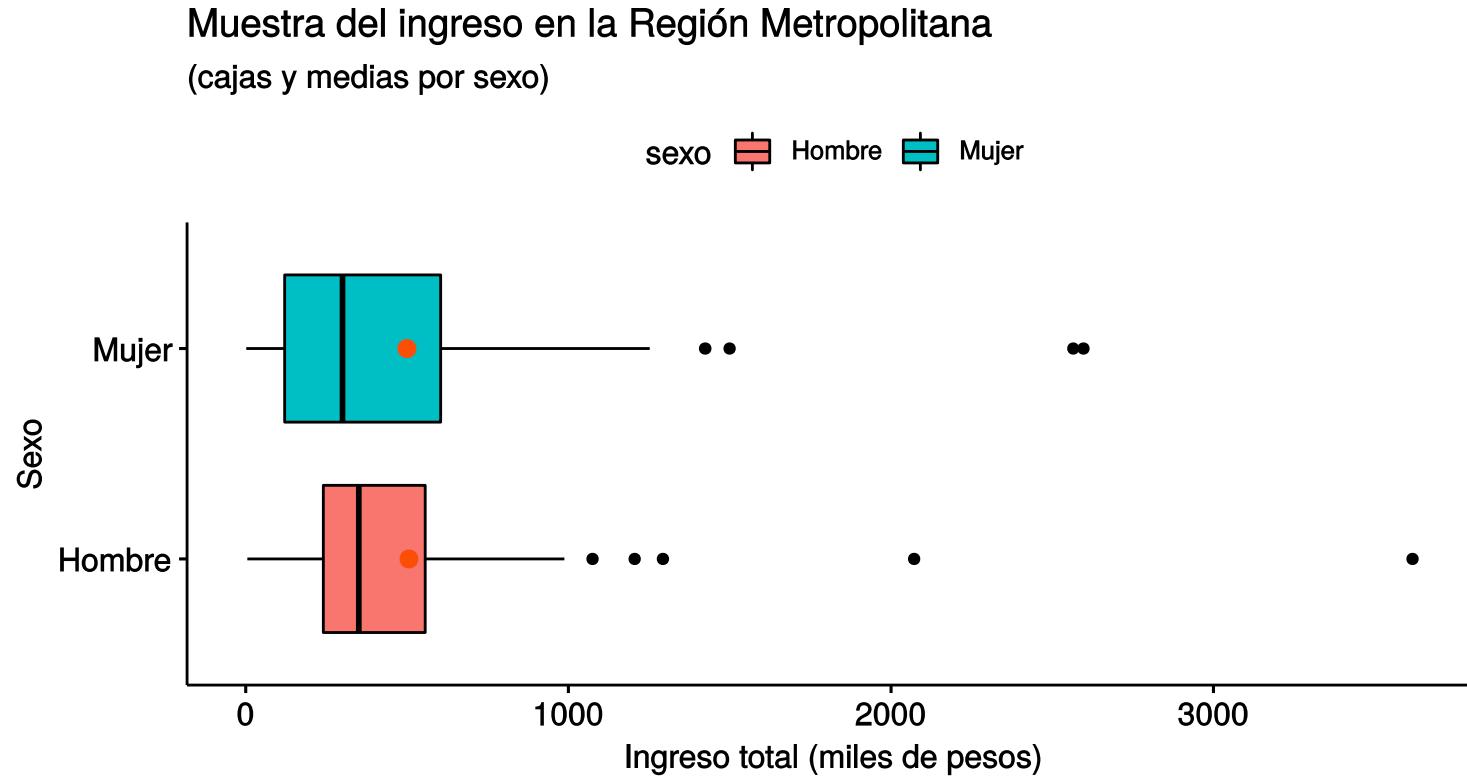
# Analizando una variable numérica en grupos distintos

- ¿Tienen hombres y mujeres **ingresos similares**?
  - nos sirven los mismos gráficos de la pregunta anterior
    - pero mostrando cada grupo **por separado**, y así
    - poder comparar la “**masa**” o/y la “**forma**” de los datos
    - basta usar la **variable de agrupación** como **color** del gráfico
    - en jerga de estadísticos, esta variable categórica se llama **factor**

```
p15 <- ggboxplot(  
  M, x = "sexo", y = "ytot",  
  add = "mean",  
  add.params = list(color = "#FC4E07", fill = "#FC4E07"),  
  fill = "sexo",  
  title = "Muestra del ingreso en la Región Metropolitana",  
  subtitle = "(cajas y medias por sexo)",  
  ylab = "Ingreso total (miles de pesos)", xlab = "Sexo"  
)  
p15 <- p15 + scale_y_continuous(limits = c(0, 3650))  
p15 <- ggpar(p15, orientation = "horizontal")
```

# Analizando una variable numérica en grupos distintos

- ¿Tienen hombres y mujeres **ingresos similares**?
  - nos sirven los mismos gráficos de la pregunta anterior
    - pero mostrando cada grupo **por separado**, y así
    - poder comparar la **“masa”** o/y la **“forma”** de los datos



# Analizando una variable numérica en grupos distintos

## ■ ¿Tienen hombres y mujeres **ingresos similares**?

- parecen tener medias similares

```
> mean(muestra[muestra[["sexo"]] == "Hombre", "ytot"])
[1] 505660.6
> mean(muestra[muestra[["sexo"]] == "Mujer", "ytot"])
[1] 499644
```

- pero ambas distribuciones son asimétricas, por lo que mejor usamos medidas más robustas

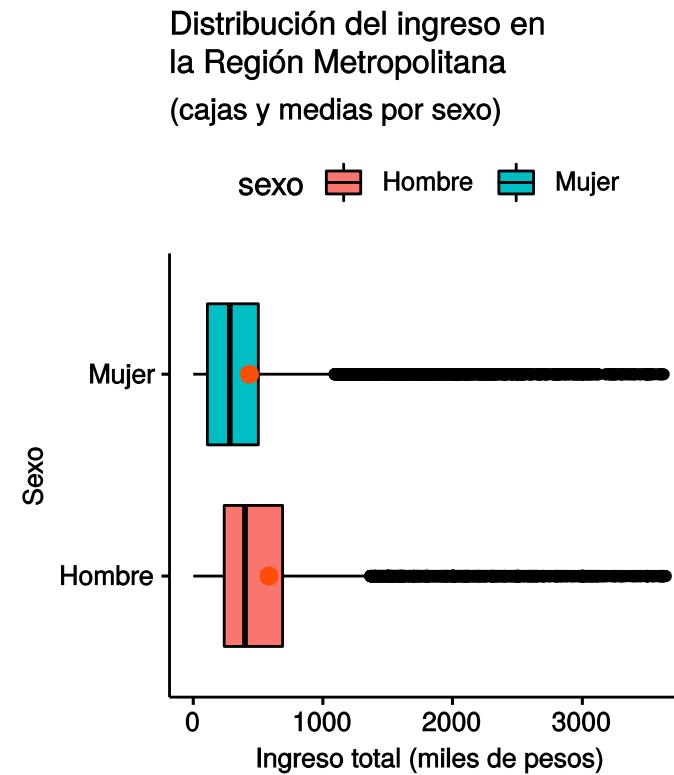
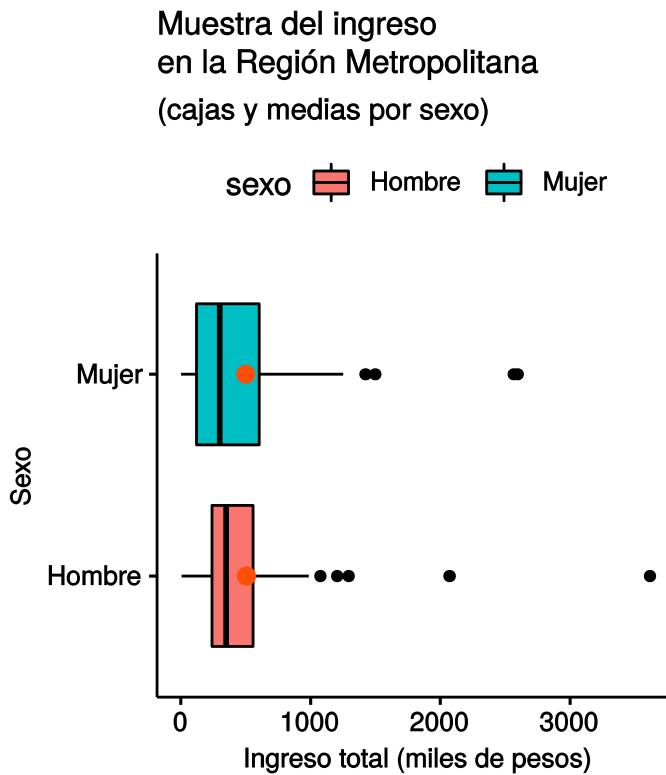
```
> median(muestra[muestra[["sexo"]] == "Hombre", "ytot"])
[1] 350000
> median(muestra[muestra[["sexo"]] == "Mujer", "ytot"])
[1] 300000
> IQR(muestra[muestra[["sexo"]] == "Hombre", "ytot"])
[1] 316250
> IQR(muestra[muestra[["sexo"]] == "Mujer", "ytot"])
[1] 483542
```

- así, el ingreso de los hombres es un poco mayor y varía menos que e de las mujeres



# Analizando una variable numérica en grupos distintos

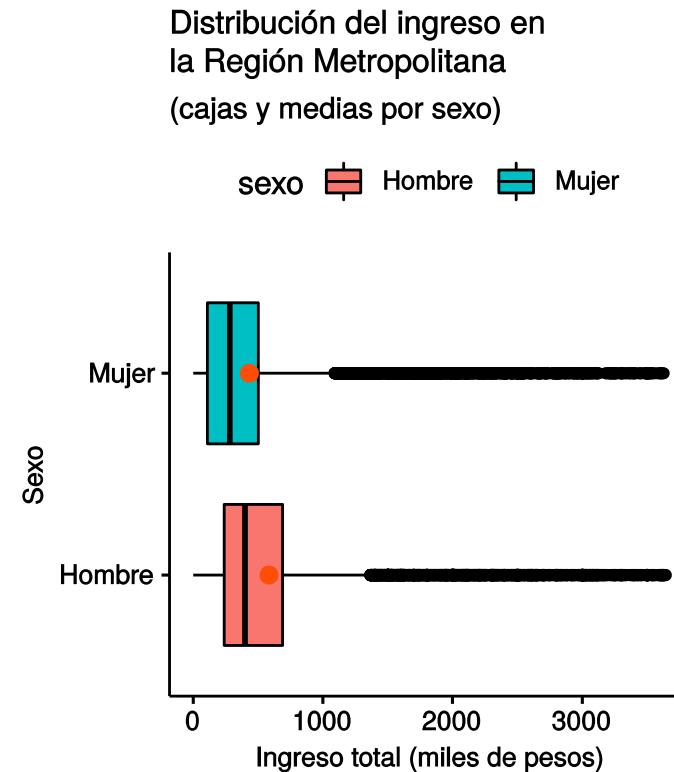
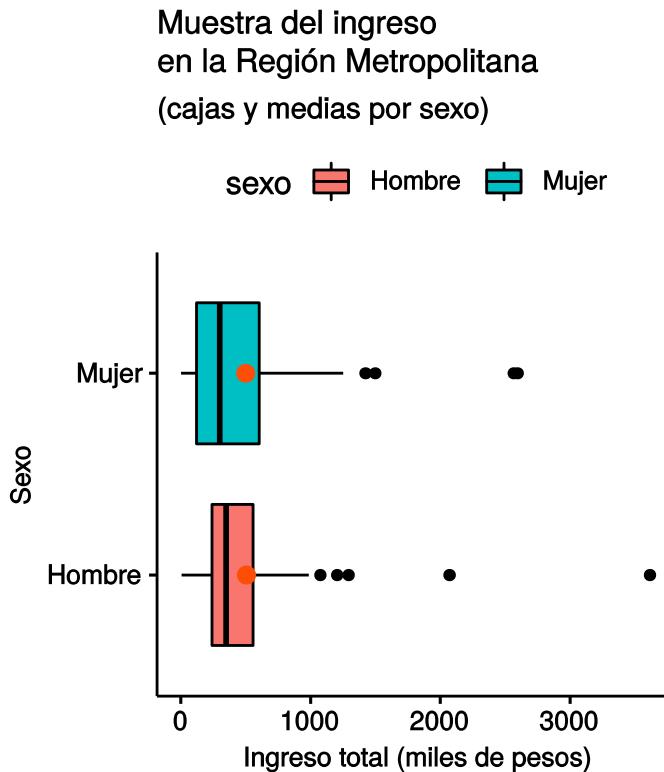
- ¿Tienen hombres y mujeres **ingresos similares**?
  - en la población, ¿se observa un comportamiento similar al observado en la muestra?





# Analizando una variable numérica en grupos distintos

- ¿Tienen hombres y mujeres **ingresos similares**?
  - en la población, ¿se observa un comportamiento similar al observado en la muestra?
    - no exactamente: los hombres parecen tener un ingreso mayor



# Analizando una variable numérica en grupos distintos

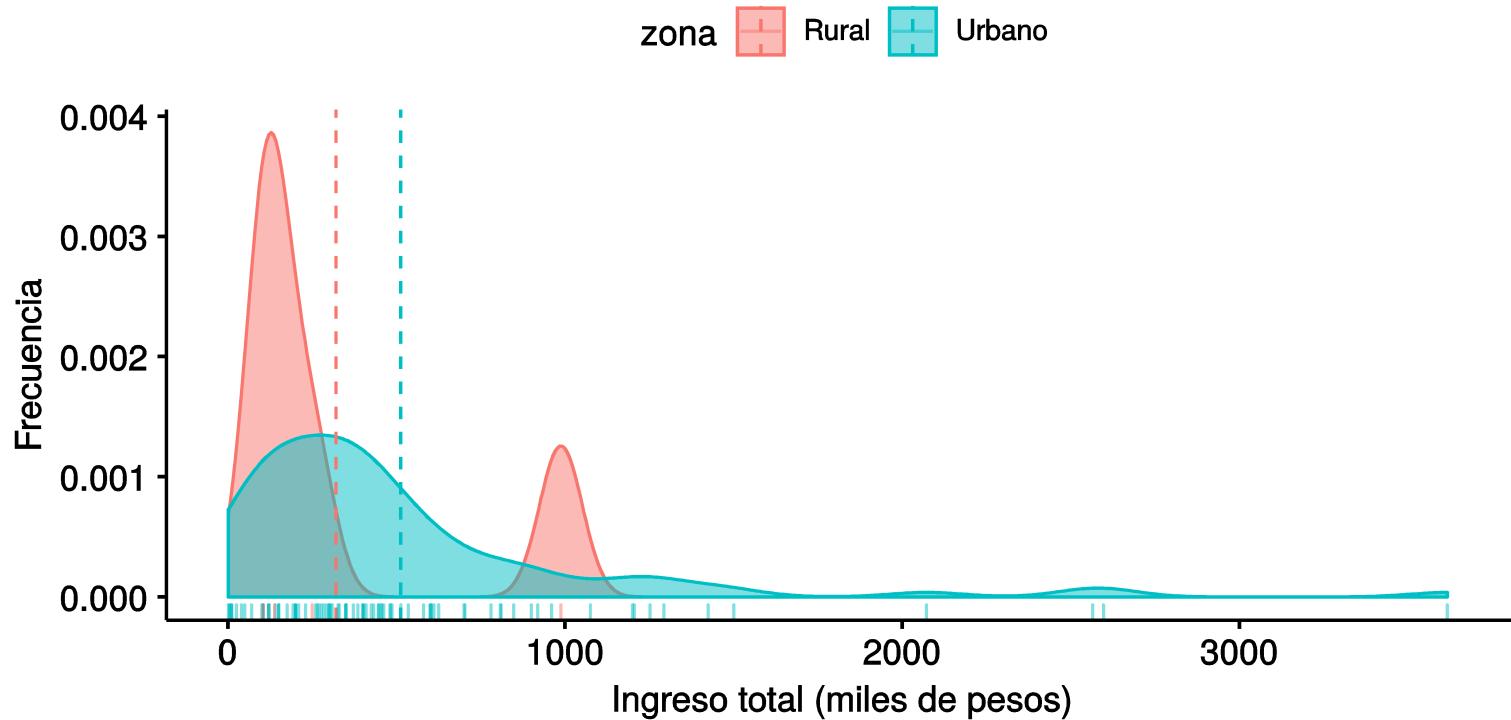
- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - mismo análisis que en la pregunta anterior

```
p17 <- ggdensity(  
  m,  
  x = "ytot",  
  add = "mean", rug = TRUE,  
  color = "zona", fill = "zona",  
  title = "Muestra del ingreso en\nla Región Metropolitana",  
  subtitle = "(distribuciones separadas por zona)",  
  xlab = "Ingreso total (miles de pesos)", ylab = "Frecuencia"  
)  
p17 <- ggpar(p17, xlim = c(0, 3650))
```

# Analizando una variable numérica en grupos distintos

- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - mismo análisis que en la pregunta anterior

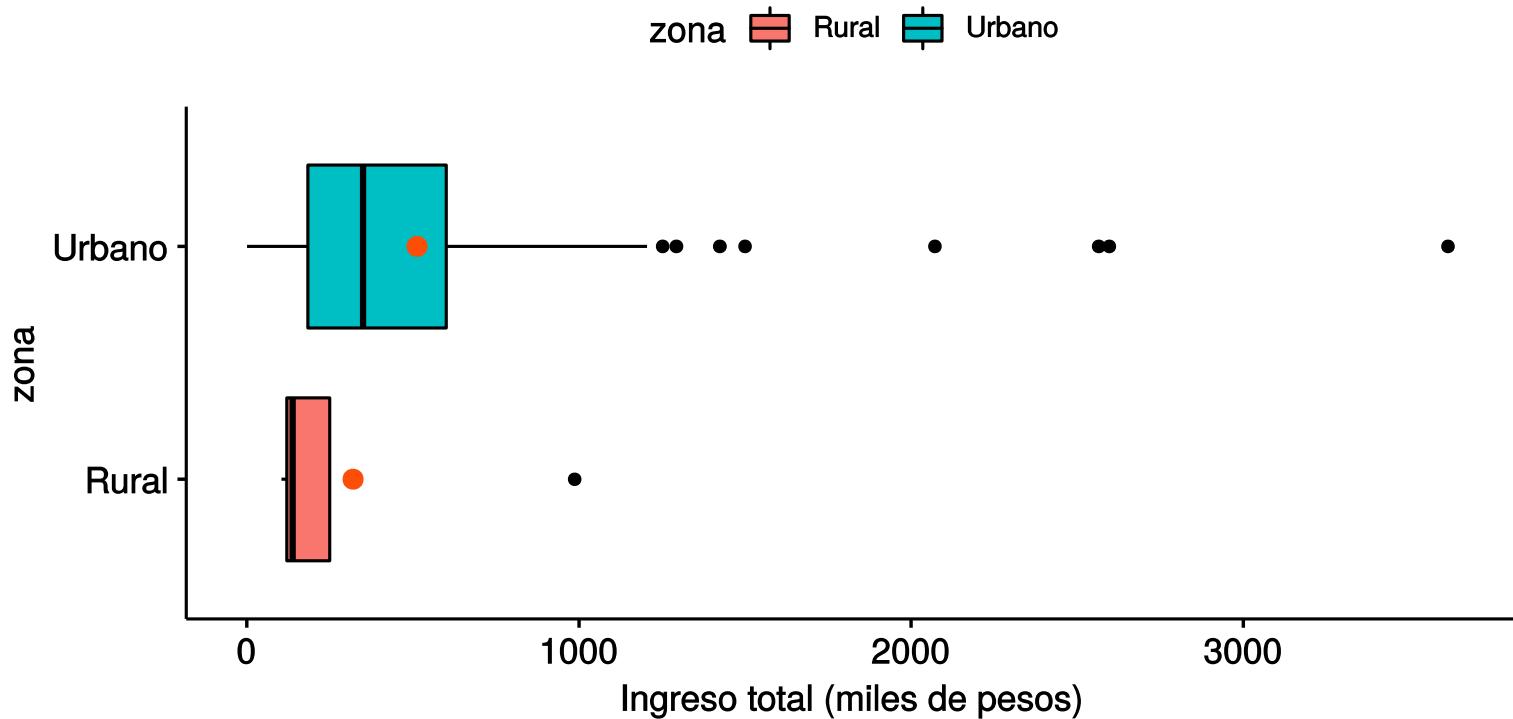
Muestra del ingreso en la Región Metropolitana  
(distribuciones separadas por zona)



# Analizando una variable numérica en grupos distintos

- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - mismo análisis que en la pregunta anterior

Muestra del ingreso en la Región Metropolitana  
(cajas y medias por zona)



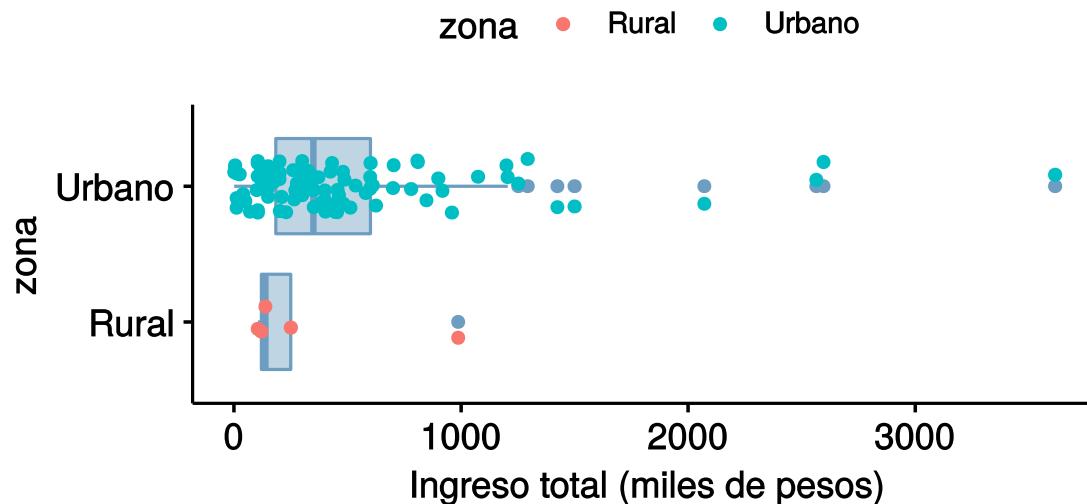
# Analizando una variable numérica en grupos distintos

- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - se aprecian **diferencias** notables entre estos grupos
  - pero, ¿podemos confiar en estos hallazgos?
  - veamos más detalle

# Analizando una variable numérica en grupos distintos

- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - se aprecian **diferencias** notables entre estos grupos
  - pero, ¿podemos confiar en estos hallazgos?
  - veamos más detalle

Muestra del ingreso en la Región Metropolitana  
(cajas y puntos por zona)



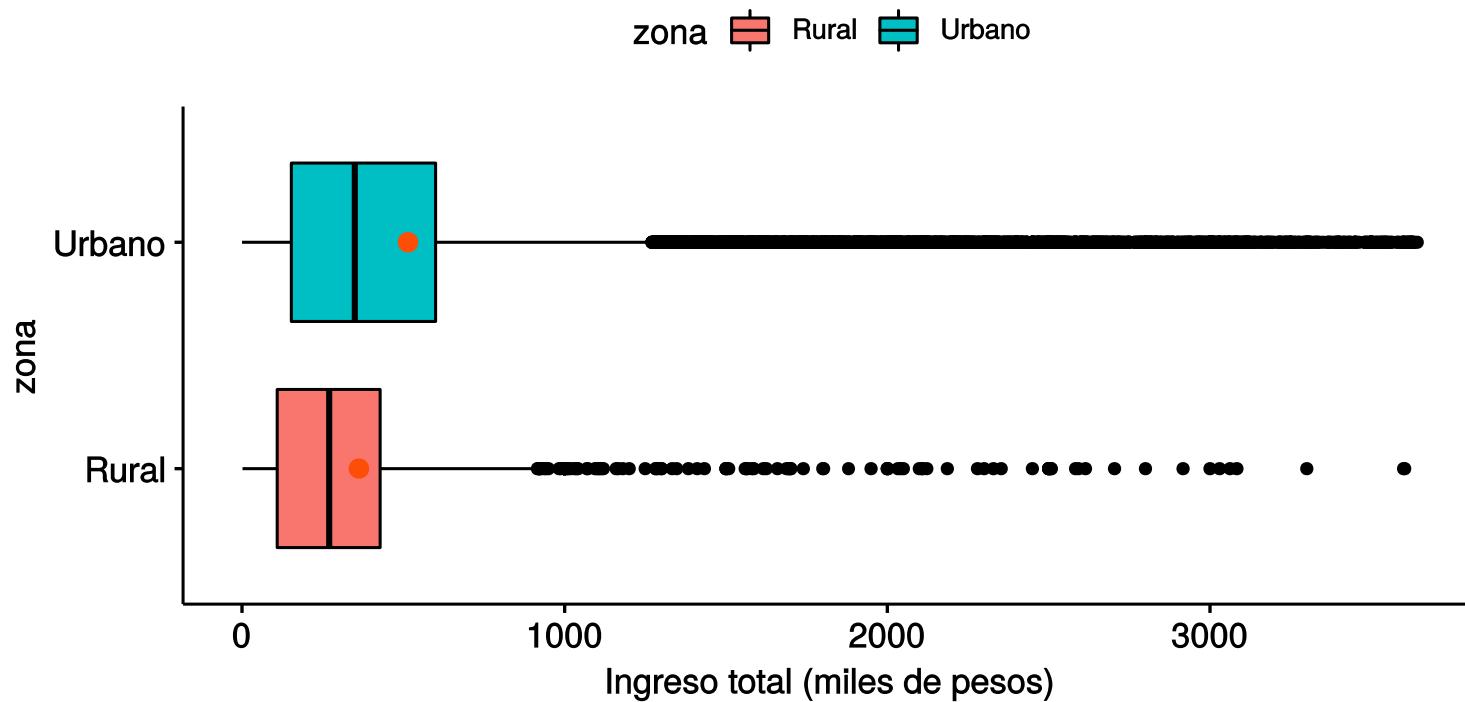
# Analizando una variable numérica en grupos distintos

- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - se aprecian **diferencias** notables entre estos grupos
  - pero, ¿podemos confiar en estos hallazgos?
  - veamos más detalle
  - desafortunadamente, la muestra capturó **muy pocos casos** de áreas rurales, por lo que no es buena idea confiar en que este comportamiento es **representativo** de la población
    - pero en este **ejercicio académico**, ¡conocemos la población!
    - veamos cómo se comparan “en realidad”

# Analizando una variable numérica en grupos distintos

- ¿Tienen las personas que viven en áreas rurales **ingresos similares** a quienes viven en áreas urbanas?
  - los ingresos parecen ser mayores en zonas urbanas

Distribución del ingreso en la Región Metropolitana  
(cajas y medias por zona)



# Analizando una variable numérica en grupos distintos

- ¿**Son similares** los ingresos registrados en las diferentes provincias de la RM?
  - mismo análisis que en la pregunta anterior, pero ahora con **más de dos grupos**

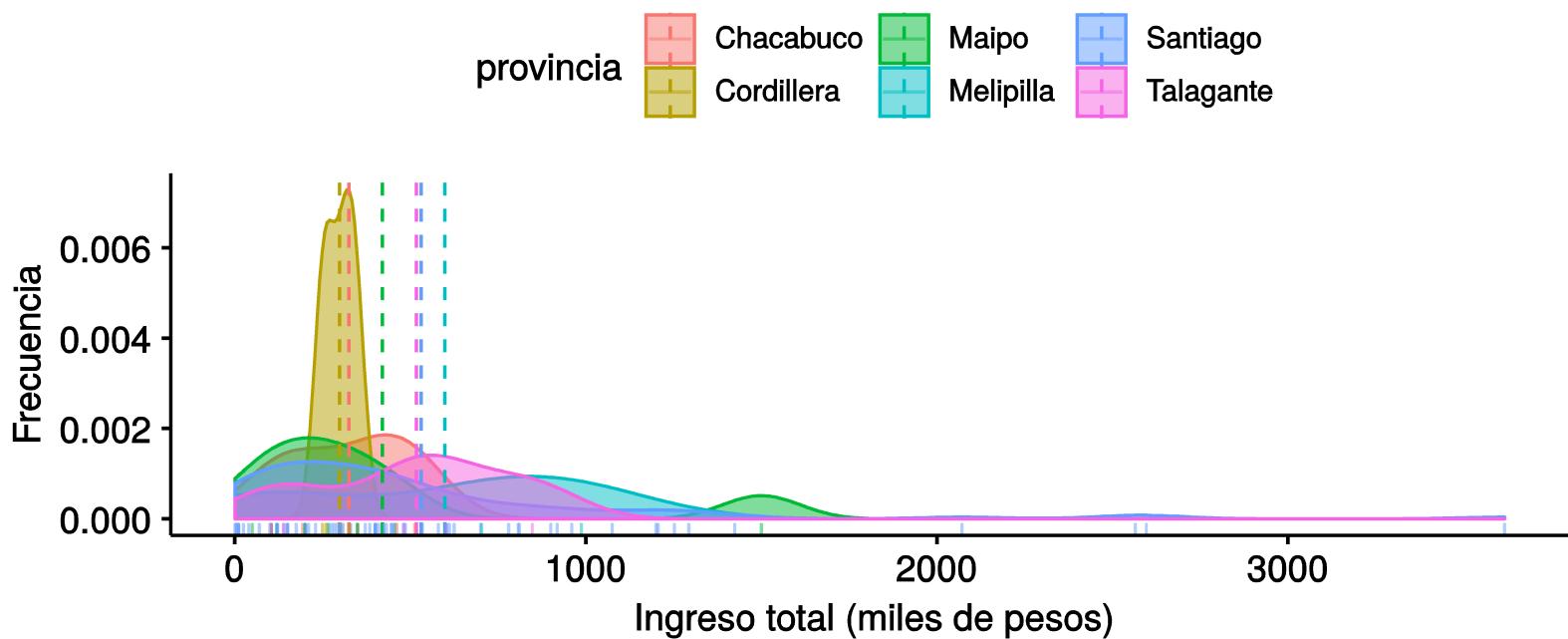
```
p22 <- ggdensity(  
  m,  
  x = "ytot",  
  add = "mean", rug = TRUE,  
  color = "provincia", fill = "provincia",  
  title = "Muestra del ingreso en la Región Metropolitana",  
  subtitle = "(distribuciones separadas por provincia)",  
  xlab = "Ingreso total (miles de pesos)", ylab = "Frecuencia"  
)  
p22 <- ggpar(p22, xlim = c(0, 3650))
```

# Analizando una variable numérica en grupos distintos

- ¿Son similares los ingresos registrados en las diferentes provincias de la RM?
  - mismo análisis que en la pregunta anterior, pero ahora con **más de dos grupos**

Muestra del ingreso en la Región Metropolitana

(distribuciones separadas por provincia)



# Analizando una variable numérica en grupos distintos

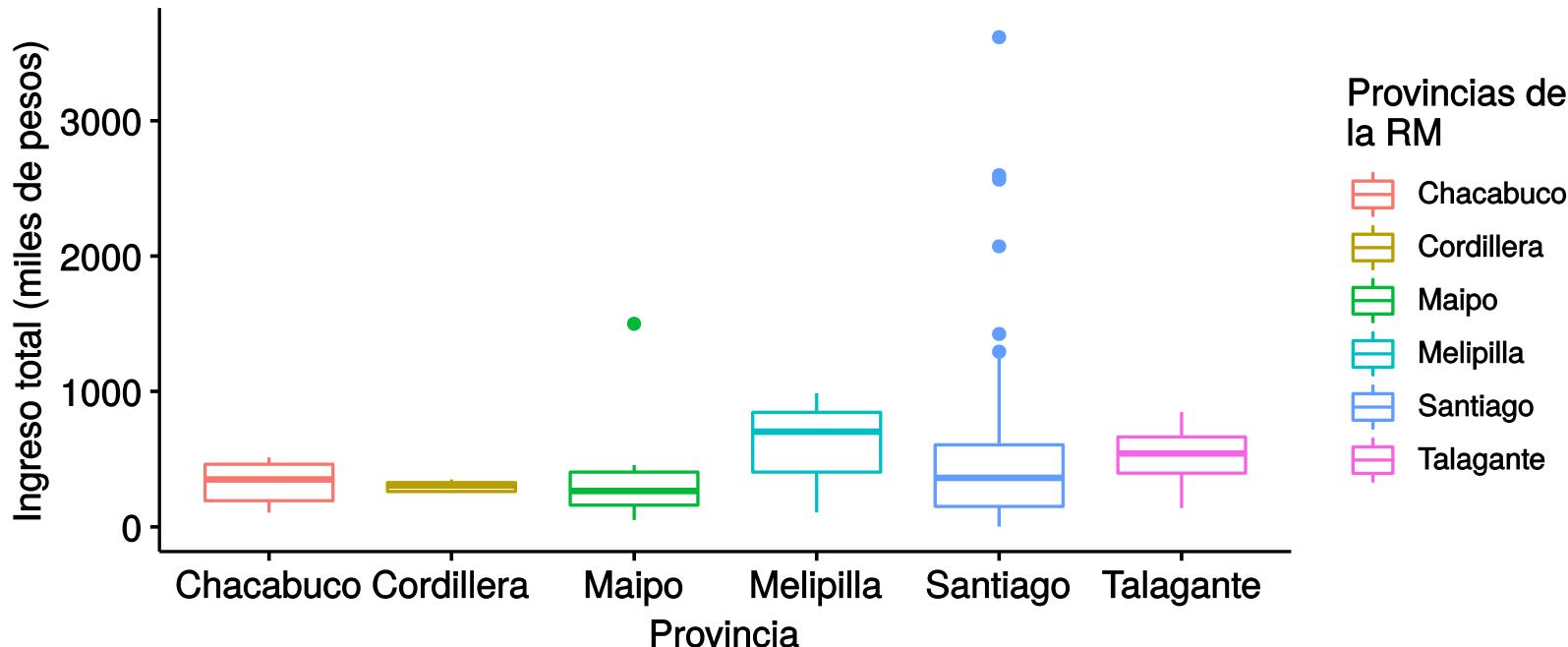
- ¿**Son similares** los ingresos registrados en las diferentes provincias de la RM?
  - mismo análisis que en la pregunta anterior, pero ahora con **más de dos grupos**

```
p24 <- ggboxplot(  
  m,  
  x = "provincia", y = "ytot",  
  color = "provincia",  
  title = "Muestra del ingreso en la Región Metropolitana",  
  subtitle = "(cajas y puntos por provincia)",  
  ylab = "Ingreso total (miles de pesos)",  
  xlab = "Provincia"  
)  
p24 <- p24 + scale_y_continuous(limits = c(0, 3650))  
p24 <- ggpar(p24, legend = "right", legend.title = "Species")  
p24 <- ggpar(p24, legend.title = "Provincias de la RM")
```

# Analizando una variable numérica en grupos distintos

- ¿Son similares los ingresos registrados en las diferentes provincias de la RM?
  - mismo análisis que en la pregunta anterior, pero ahora con **más de dos grupos**

Muestra del ingreso en la Región Metropolitana  
(cajas y puntos por provincia)



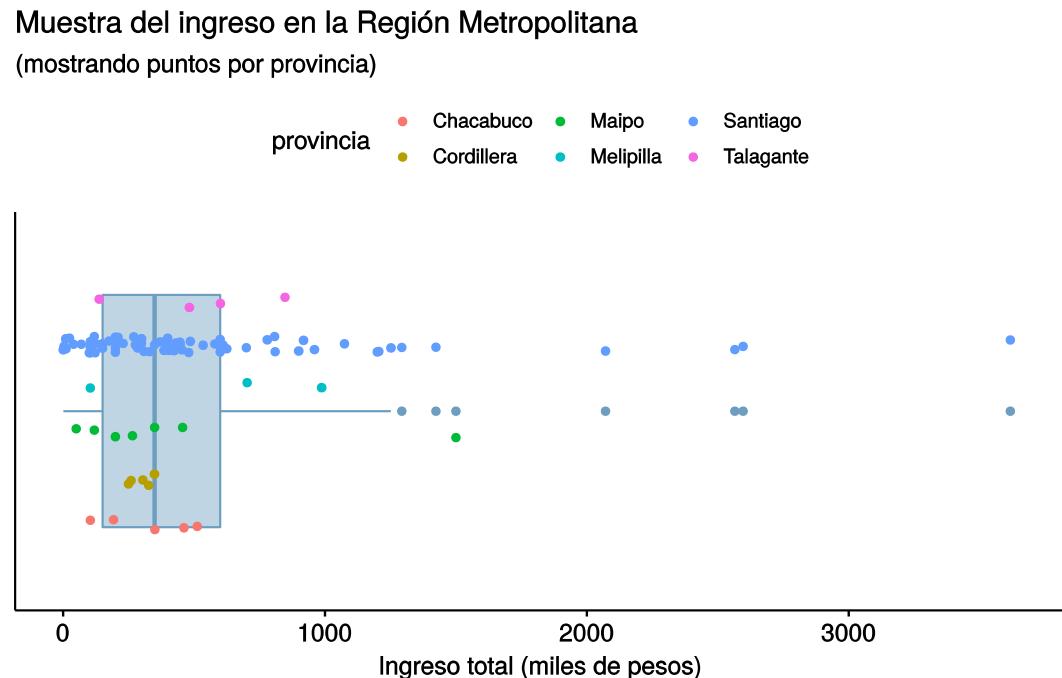
# Analizando una variable numérica en grupos distintos

- ¿**Son similares** los ingresos registrados en las diferentes provincias de la RM?
  - mismo análisis que en la pregunta anterior, pero ahora con **más de dos grupos**
  - pero, también debemos verificar la validez de la muestra



# Analizando una variable numérica en grupos distintos

- ¿Son similares los ingresos registrados en las diferentes provincias de la RM?
  - mismo análisis que en la pregunta anterior, pero ahora con **más de dos grupos**
  - pero, también debemos verificar la validez de la muestra
  - vemos que hay grupos con 3 y 4 puntos
  - por lo que es **arriesgado** hacer *inferencias* con esta muestra



# Analizando dos variables numéricas

- ¿Tiene relación el ingreso con la riqueza del municipio donde se habita?
  - aquí se nos pregunta por la **influencia** de una **variable numérica** en **otra variable numérica**
    - si una crece, ¿la otra también?
    - o al revés, si la primera crece, ¿la otra disminuye?
    - o simplemente no hay relación (independencia)
  - un **gráfico de dispersión**, y un buen ojo, suele ser adecuado en estos casos

# Analizando dos variables numéricas

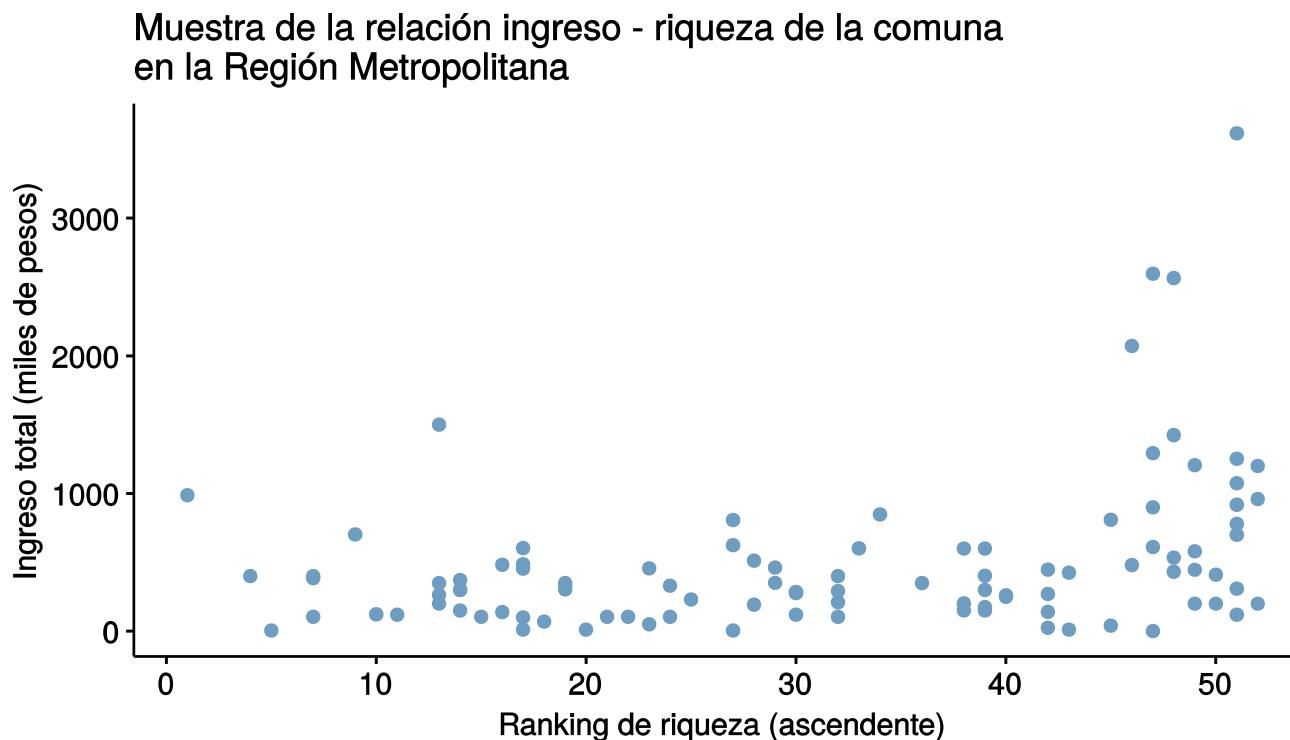
- ¿Tiene relación el ingreso con la riqueza del municipio donde se habita?

```
p29 <- ggscatter(  
  m,  
  x = "ing.comuna", y = "ytot",  
  color = "#6D9EC1", fill = "#BFD5E3",  
  title = "Muestra de la relación ingreso - riqueza de la comuna\nen la Región  
Metropolitana",  
  ylab = "Ingreso total (miles de pesos)",  
  xlab = "Ranking de riqueza (ascendente)"  
)  
p29 <- p29 + scale_y_continuous(limits = c(0, 3650))
```



# Analizando dos variables numéricas

- ¿Tiene relación el ingreso con la riqueza del municipio donde se habita?
  - ¿parece haber una relación algo positiva?



# Analizando dos variables numéricas

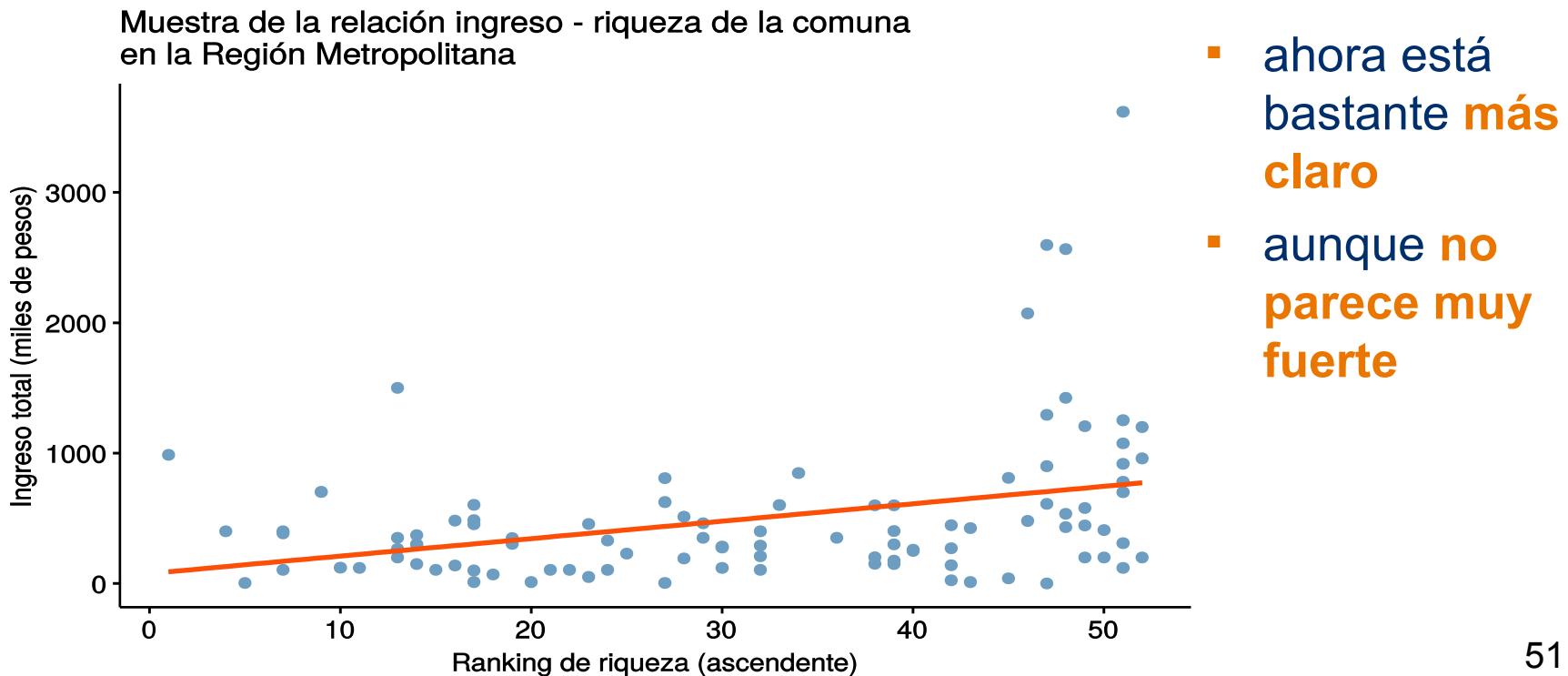
- ¿Tiene relación el ingreso con la riqueza del municipio donde se habita?
  - ¿parece haber una relación algo positiva?
  - a veces ayuda agregar una Línea con la tendencia

```
p29 <- ggscatter(  
  m,  
  x = "ing.comuna", y = "ytot",  
  add = "reg.line", add.params = list(color = "#FC4E07"),  
  color = "#6D9EC1", fill = "#BFD5E3",  
  title = "Muestra de la relación ingreso - riqueza de la comuna\nen la Región  
Metropolitana",  
  ylab = "Ingreso total (miles de pesos)",  
  xlab = "Ranking de riqueza (ascendente)"  
)  
p29 <- p29 + scale_y_continuous(limits = c(0, 3650))
```



# Analizando dos variables numéricas

- ¿Tiene relación el ingreso con la riqueza del municipio donde se habita?
  - ¿parece haber una relación algo positiva?
  - a veces ayuda agregar una línea con la tendencia

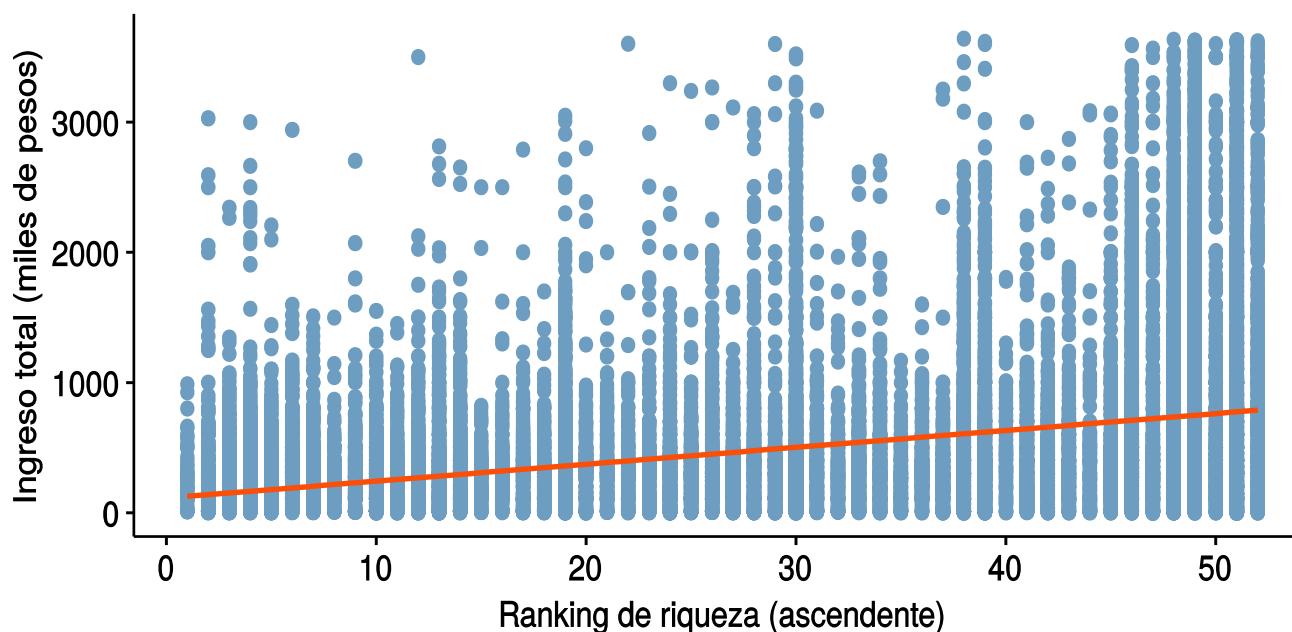




# Analizando dos variables numéricas

- ¿Tiene relación el ingreso con la riqueza del municipio donde se habita?
  - ¿se observa esta relación en la población?
    - parece que bastante bien

Muestra de la relación ingreso - riqueza de la comuna  
en la Región Metropolitana



# Analizando dos variables numéricas en grupos distintos

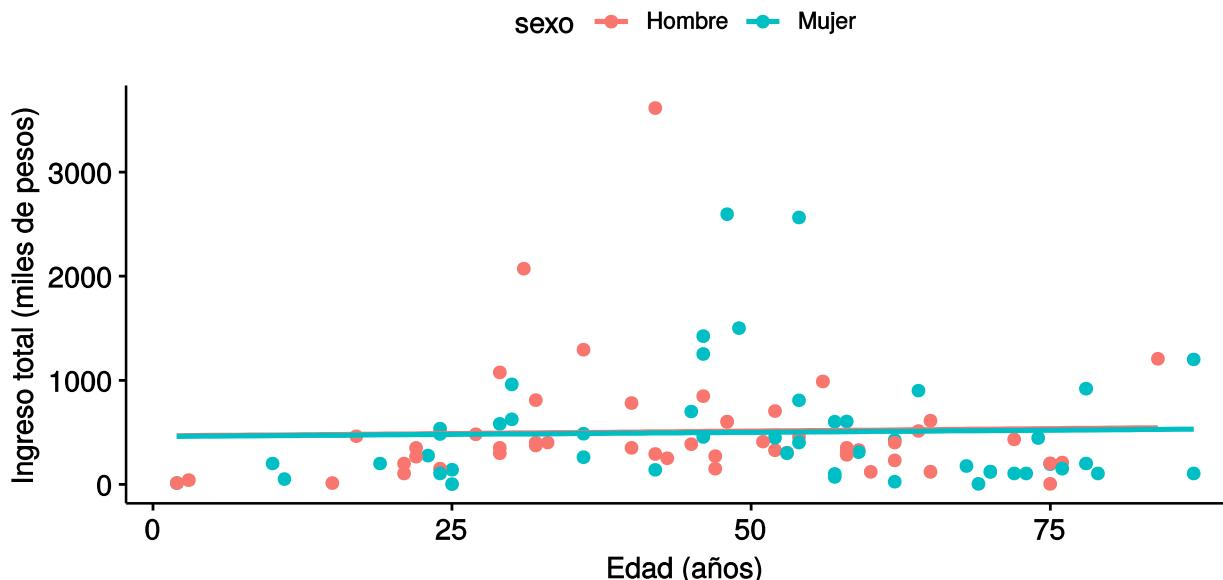
- ¿Van los ingresos de los chilenos **incrementándose** con la experiencia y **de forma similar entre hombres y mujeres?**
  - análisis muy similar, pero ahora sepáramos grupos
    - con colores

```
p28 <- ggscatter(  
  m,  
  x = "edad", y = "ytot",  
  add = "reg.line", add.params = list(color = "sexo"),  
  color = "sexo", fill = "sexo",  
  title = "Relación ingreso - edad\nen la Región Metropolitana",  
  subtitle = "(separada por sexo)",  
  ylab = "Ingreso total (miles de pesos)",  
  xlab = "Edad (años)"  
)  
p28 <- p28 + scale_y_continuous(limits = c(0, 3650))
```

# Analizando dos variables numéricas en grupos distintos

- ¿Van los ingresos de los chilenos **incrementándose** con la experiencia y **de forma similar entre hombres y mujeres?**
  - análisis muy similar, pero ahora separamos grupos
    - con colores

Muestra de la relación ingreso - edad en la Región Metropolitana  
(separada por sexo)



# Analizando dos variables numéricas en grupos distintos

- ¿Van los ingresos de los chilenos **incrementándose** con la experiencia y **de forma similar entre hombres y mujeres**?
  - análisis muy similar, pero ahora separamos grupos
    - con colores
  - parece que **no hay relación**, en **ninguno** de los grupos
    - por lo que la respuesta a la pregunta sería “no, los ingresos no aumentan con la experiencia, ni para hombres ni para mujeres”
    - ¿triste?
    - ¿seguirá este comportamiento en la población?

# Analizando dos variables numéricas en grupos distintos

- ¿Van los ingresos de los chilenos **incrementándose** con la experiencia y **de forma similar** entre hombres y mujeres?
  - se mantiene ¡para las mujeres!

Muestra de la relación ingreso - edad en la Región Metropolitana  
(separada por sexo)

