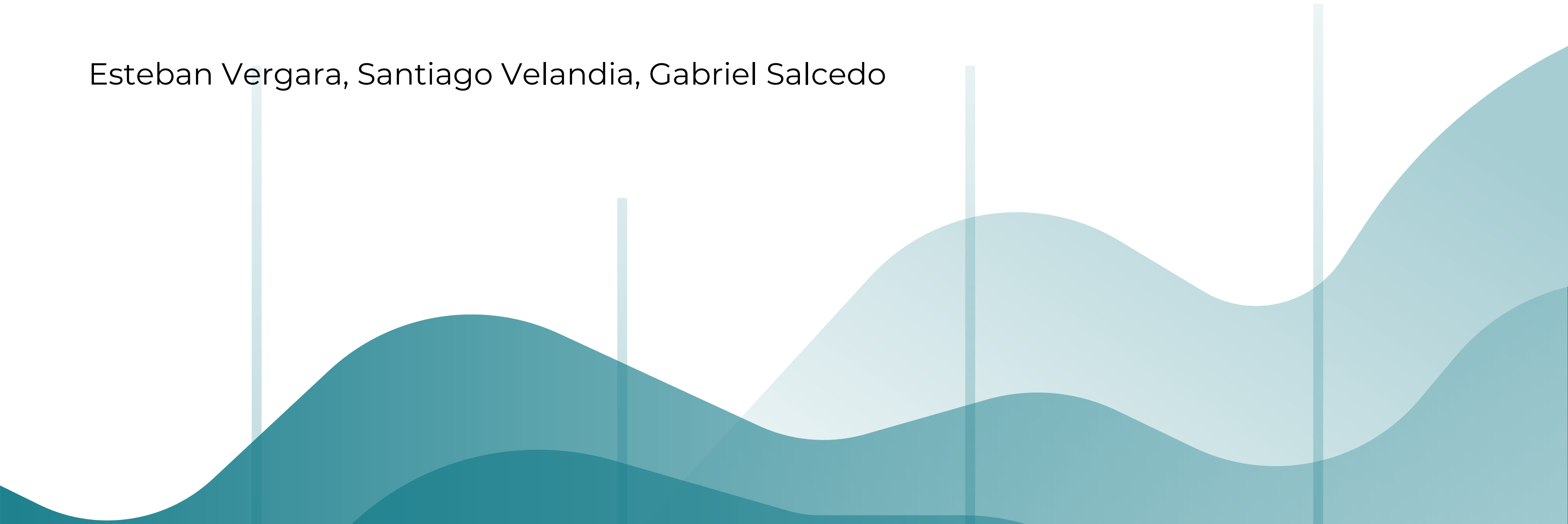


TALLER 1: IDENTIFICANDO FUNCIONES BÁSICAS

¿Vale la pena invertir en capacitación para tenderos?

Esteban Vergara, Santiago Velandia, Gabriel Salcedo



Objetivo de la Investigación



Identificar y recopilar datos de una variable, para realizar una gráfica e interpretar los resultados obtenidos.



Variable utilizada

702. En los últimos 2 años ¿usted o alguien que participe en la toma de decisiones ha recibido algún tipo de capacitación para la actividad económica que realiza?

SELECCIÓN-ÚNICA

Capac

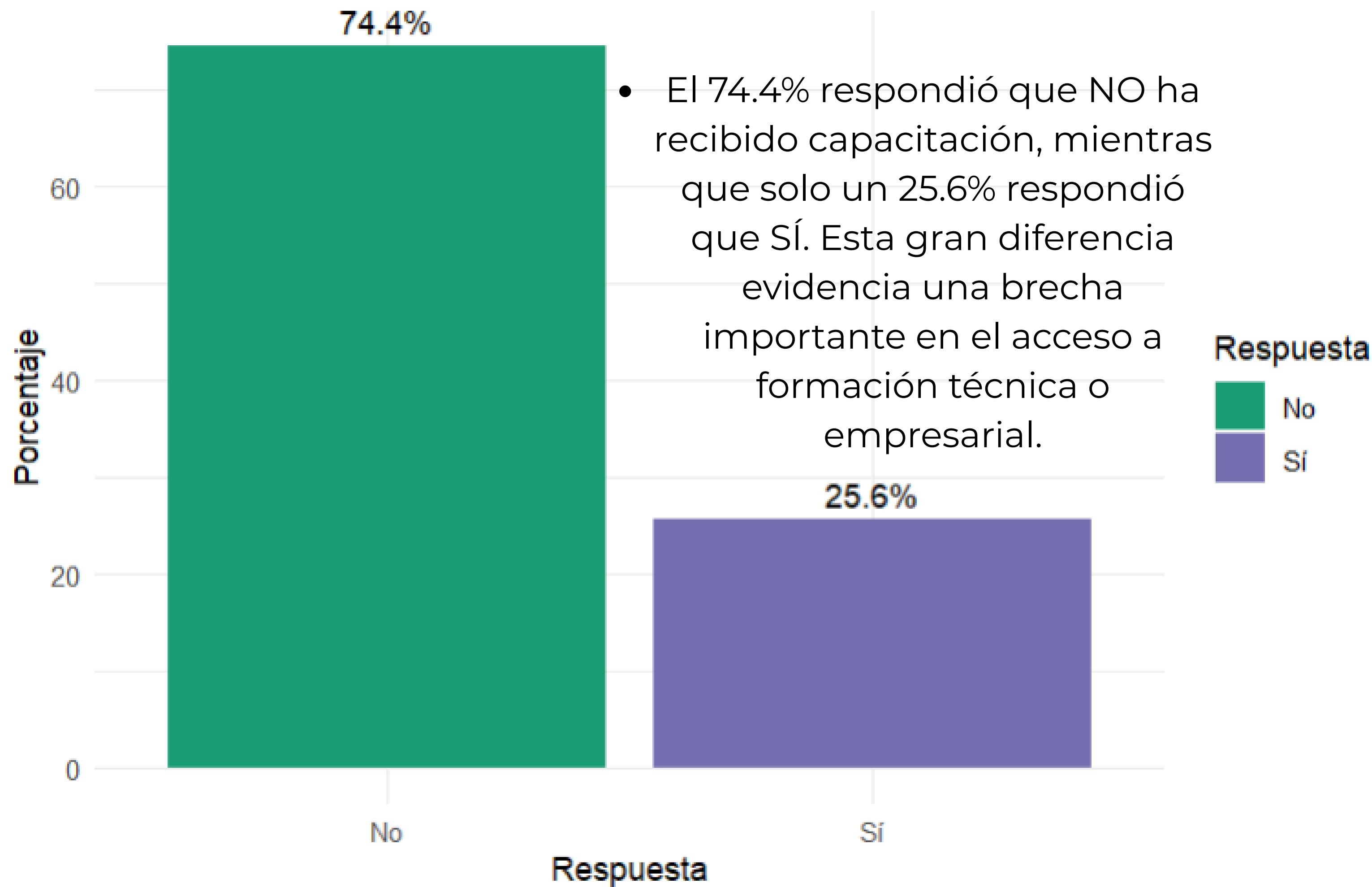
01 ☐ Sí

00 ☐ No

Organización del Código

```
1  # Instala ggplot2 solo la primera vez (luego puedes comentar esta línea)
2  install.packages("ggplot2")
3
4  # Carga la librería de gráficos
5  library(ggplot2)
6
7  # Observa la frecuencia de cada respuesta en la variable 'Capac' del dataframe principal
8  table(TenderosFU03_Publica$Capac)
9
10 # Calcula el porcentaje de cada respuesta
11 prop.table(table(TenderosFU03_Publica$Capac)) * 100
12
13 # Crea un resumen (dataframe) con los porcentajes obtenidos anteriormente
14 resumen <- data.frame(
15   Respuesta = c("No", "Sí"),
16   Porcentaje = c(74.42, 25.57)
17 )
18
19 # Muestra el resumen para verificar su contenido
20 print(resumen)
21
22 # Grafica los resultados con ggplot2
23 ggplot(resumen, aes(x = Respuesta, y = Porcentaje, fill = Respuesta)) +
24   geom_bar(stat = "identity") + # Dibuja las barras según los porcentajes
25   geom_text(aes(label = paste0(round(Porcentaje, 1), "%")), vjust = -0.5) + # Agrega etiquetas de porcentaje arriba de cada barra
26   scale_fill_manual(values = c("#1b9e77", "#7570b3")) + # Colores de la gráfica
27   labs(
28     title = "¿En los últimos 2 años... ha recibido capacitación?",
29     x = "Respuesta",
30     y = "Porcentaje"
31   ) +
32   theme_minimal() # Usa un tema visual limpio
```

¿En los últimos 2 años... ha recibido capacitación?



¿Vale la pena invertir en capacitación para tenderos?

De acuerdo con los resultados, únicamente el 25.6% de los encuestados manifestó haber recibido algún tipo de capacitación, en contraste con el 74.4% que indicó no haber accedido a este tipo de formación. Esta disparidad pone de manifiesto una brecha considerable en el acceso a programas de capacitación técnica o empresarial, lo cual podría tener implicaciones relevantes en el desarrollo de competencias y oportunidades laborales.





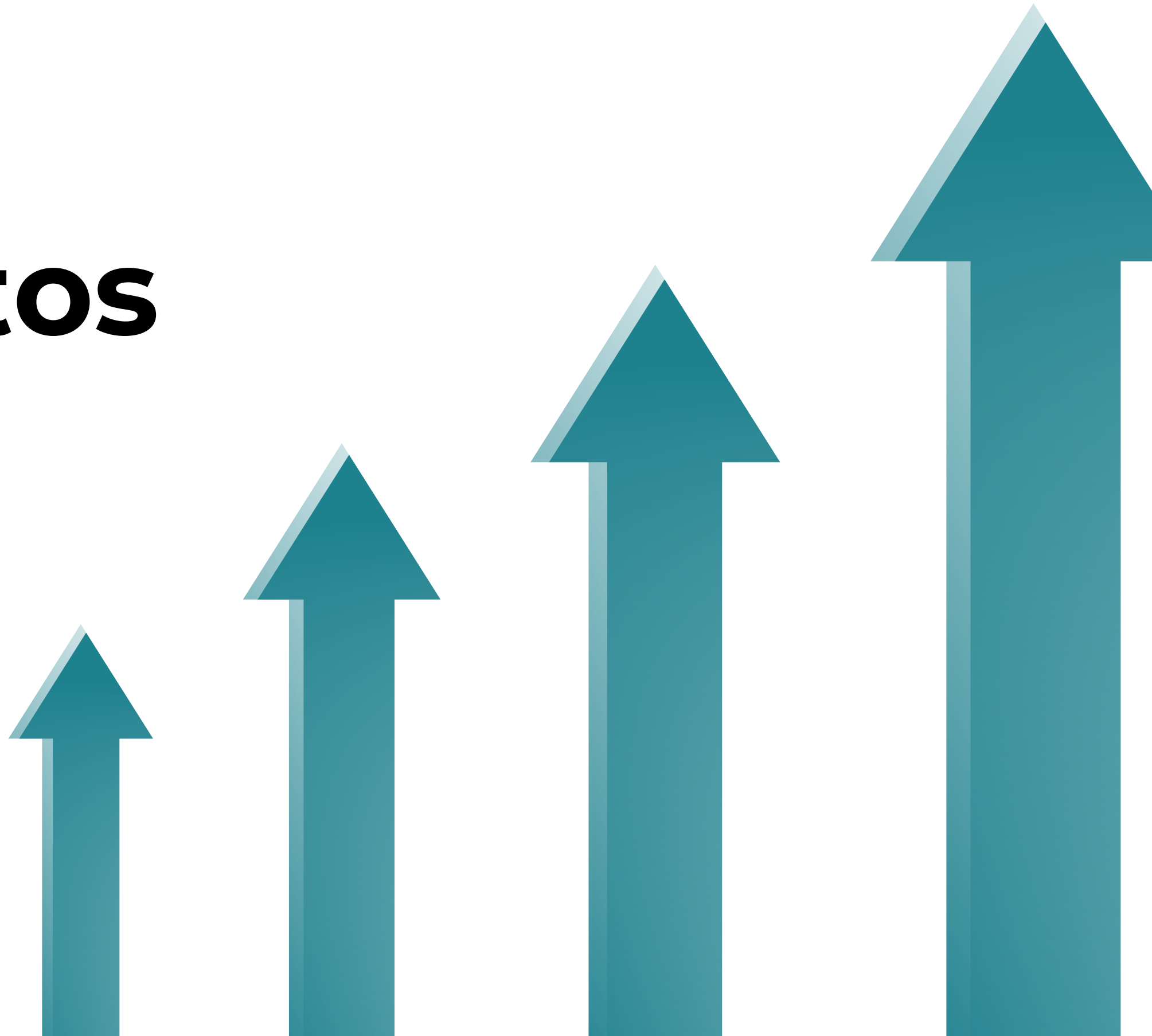
Conclusión

Los datos muestran una brecha significativa en el acceso a la capacitación técnica o empresarial, con solo el 25.6% de los encuestados afirmando haber recibido formación. Esta diferencia refleja no solo limitaciones estructurales como la desigualdad o la escasa oferta institucional, sino también factores individuales como la motivación y la disposición a aprender.

Por lo tanto, es necesario implementar políticas públicas que no solo amplíen el acceso a la formación, sino que también promuevan el interés y la participación activa de la población, reconociendo la capacitación como un pilar para el desarrollo económico y social sostenible.

Taller 2:

Manejo de datos





Desarrollo del código

Utilizamos las librerías: dplyr, tidyr para manejar los datos. La librería scales, para aproximar valores porcentuales y la librería writexl para exportar los dataframes obtenidos como archivos xlsx para luego realizar los gráficos en PowerBI.

La primera parte del código carga los archivos y hace un primer filtro en la variables 'uso_internet' del archivo 'Tenderos' para asegurarnos de que esta tenga valores con los que podamos trabajar. Posterior a esto hacemos nuestro primer dataframe con el promedio de internet por municipio.

```
# Cargamos las librerías necesarias para manipulación de datos
library(dplyr)
library(tidyr)
library(scales)

#Paths datasets utilizados

TenderosFU03_Publica <- read_dta("D:/universidad/2 semestre/haciendo economía/TenderosFU03_Publica.dta")
TerriData_Dim2_Sub3 <- read_excel("D:/universidad/2 semestre/haciendo economía/TerriData_Dim2_Sub3.xlsx")

#Elimina las etiquetas y deja los valores subyacentes como numéricos
TenderosFU03_Publica <- TenderosFU03_Publica %>%
  mutate(uso_internet = zap_labels(uso_internet))

# -----
# TAREA 1: Promedio de uso de internet por municipio
# -----

internet_munic <- TenderosFU03_Publica %>%
  group_by(Munic_Dept, Municipio) %>% # Agrupa los datos por departamento y municipio
  summarize(internet = mean(uso_internet, na.rm = TRUE) * 100) %>% # Calcula el promedio de uso de internet (en porcentaje)
  ungroup() %>% # Desagrupa los datos
  rename(divipola = Munic_Dept) # Renombra la columna de departamento a "divipola"
```



Desarrollo del código

En la segunda parte de nuestro código teníamos como objetivo obtener los datos de actividad económica, para conseguir esto utilizamos la función 'pivot_longer' con el fin de pasar de tener los datos en diferentes columnas a una sola.

Antes de hacer eso, renombramos cada variable con el fin de facilitar la identificación de cada una de ellas, en el momento de realizar el 'pivot_longer'.

Finalmente, obtenemos un dataframe con el uso promedio de internet según actividad económica.

```

25 # -----
26 # TAREA 2: Transformar datos de actividades y calcular uso de internet por tipo de actividad
27 # -----
28
29 #Limpiamos y organizamos datos de tal forma que se puedan agrupar por actividad
30
31 ren_frame <- TenderosFU03_Publica %>%
32   # Renombramos las columnas de actividades con nombres mas descriptivos
33   rename(Tienda.1 = actG1, ComidaPreparada.2 = actG2, Peluqueria.3 = actG3, Ropa.4 = actG4,
34          Otras.5 = actG5, Papeleria.6 = actG6, vidanocturna.7 = actG7, productosinventario.8 = actG8,
35          salud.9 = actG9, servicios.10 = actG10, ferreteria.11 = actG11)
36
37 col_frame <- ren_frame %>%
38   # Convertimos las columnas de actividades en dos columnas: "category" (nombre original) y "total" (valor 0/1)
39   pivot_longer(cols = Tienda.1:ferreteria.11, names_to = "category", values_to = "total")
40
41 col_frame <- col_frame %>%
42   # Separamos la columna "category" en dos: el nombre de la actividad y su numero
43   separate(category, c("Actividad", "actG"))
44
45 #Frame con uso de internet por actividad economica
46
47 internet_act <- col_frame %>%
48   group_by(actG, Actividad) %>% # Agrupamos por numero de actividad y nombre
49   summarize(internet = mean(uso_internet, na.rm = TRUE) * 100) # Calculamos el promedio de uso de internet por actividad
50

```



Desarrollo del código

Para la tercera parte del código utilizamos lo que ya hemos hecho anteriormente para crear un dataframe que nos muestre el promedio de uso de internet por municipio y por actividad económica.

Este dataframe será fundamental para dar respuesta a una de las preguntas planteadas. Adicionalmente, será utilizado después de procesar la información demográfica.

```
# -----  
# TAREA 3: Promedio de uso de internet por municipio y actividad (solo donde la actividad esta presente)  
# -----  
  
internet_act_munic <- col_frame %>%  
  filter(total == 1) %>% # Filtramos solo las filas donde la actividad está presente (valor 1)  
  group_by(Munic_Dept, Municipio, actG, Actividad) %>% # Agrupamos por municipio y tipo de actividad  
  summarise(internet = percent(mean(uso_internet, na.rm = TRUE), accuracy = 0.1)) %>% # Calculamos el promedio de uso de internet  
  rename(divipola = Munic_Dept) %>% # Renombramos la columna del departamento  
  arrange(Municipio, actG) # Ordenamos por municipio y tipo de actividad
```



Desarrollo del código

La cuarta parte de nuestro código procesa la información del archivo 'Terridata'. Para esto eliminamos la primera fila del archivo, puesto que esta nos generaba un error. Posterior a esto creamos dos nuevas columnas, las cuales limpiaron los 'Datos Numéricos' y el 'Codigo' (divipola), puesto eran datos numéricos incorrectos que podrían generar errores.

Después de este proceso de limpieza, creamos el dataframe con la información demográfica de cada municipio (divipola).

```

62 # -----
63 # TAREA 4: Procesar datos demograficos para obtener poblacion total por municipio
64 # -----
65
66 #Limpieza de base datos
67
68 TerriData_Dim2_Sub3 <- TerriData_Dim2_Sub3[-1,]
69
70 TerriData_Dim2_Sub3 <- TerriData_Dim2_Sub3 %>%
71   mutate(
72     `Dato Numérico Limpio` = `Dato Numérico` %>%
73       gsub("\\\\.", "", .) %>%           # Quitar puntos de miles
74       gsub(",00$", "", .) %>%          # Quitar ",00" si está al final
75       gsub(",", ".", .) %>%           # Si hay otros decimales, convertir a punto decimal
76       as.numeric()                     # Convertir a número
77   )
78
79 TerriData_Dim2_Sub3 <- TerriData_Dim2_Sub3 %>%
80   mutate(divipola=as.numeric(`Código Entidad`))
81
82 #Frame con poblacion por codigo DANE
83
84 poblacion <- TerriData_Dim2_Sub3 %>%
85   group_by(divipola) %>%
86   filter(Año==2024, `Unidad de Medida`!= "Porcentaje (el valor está multiplicado por 100)")%>%
87   summarise(poblacion = sum(`Dato Numérico Limpio`))
88

```




Desarrollo del código

La última parte de nuestro código fue realizar un full merge entre el dataframe creado en la parte tres de nuestro código y el dataframe obtenido en paso anterior. Es decir que en este dataframe tenemos el uso promedio de internet por municipio y por actividad, y el código divipola y la población de cada municipio.

Además exportamos este dataframe (final_dataframe) y el tercero (internet_act_munic) como archivos xlsx para graficarlos en PowerBI.

```
90 # -----
91 # TAREA 5: Utilizamos merge() para unir las dos bases de datos principales
92 # -----
93
94 final_dataframe <- merge(poblacion, internet_act_munic, by.x = "divipola", by.y = "divipola")
95
96 # -----
97 #Exportamos dataframes a excel para graficar con powerbi utilizando libreria writexl
98 # -----
99
100 library(writexl)
101
102 # Guardar como Excel cada dataframe
103
104 write_xlsx(internet_act_munic, "internet_act_munic.xlsx")
105 write_xlsx(final_dataframe, "final_dataframe.xlsx")
```

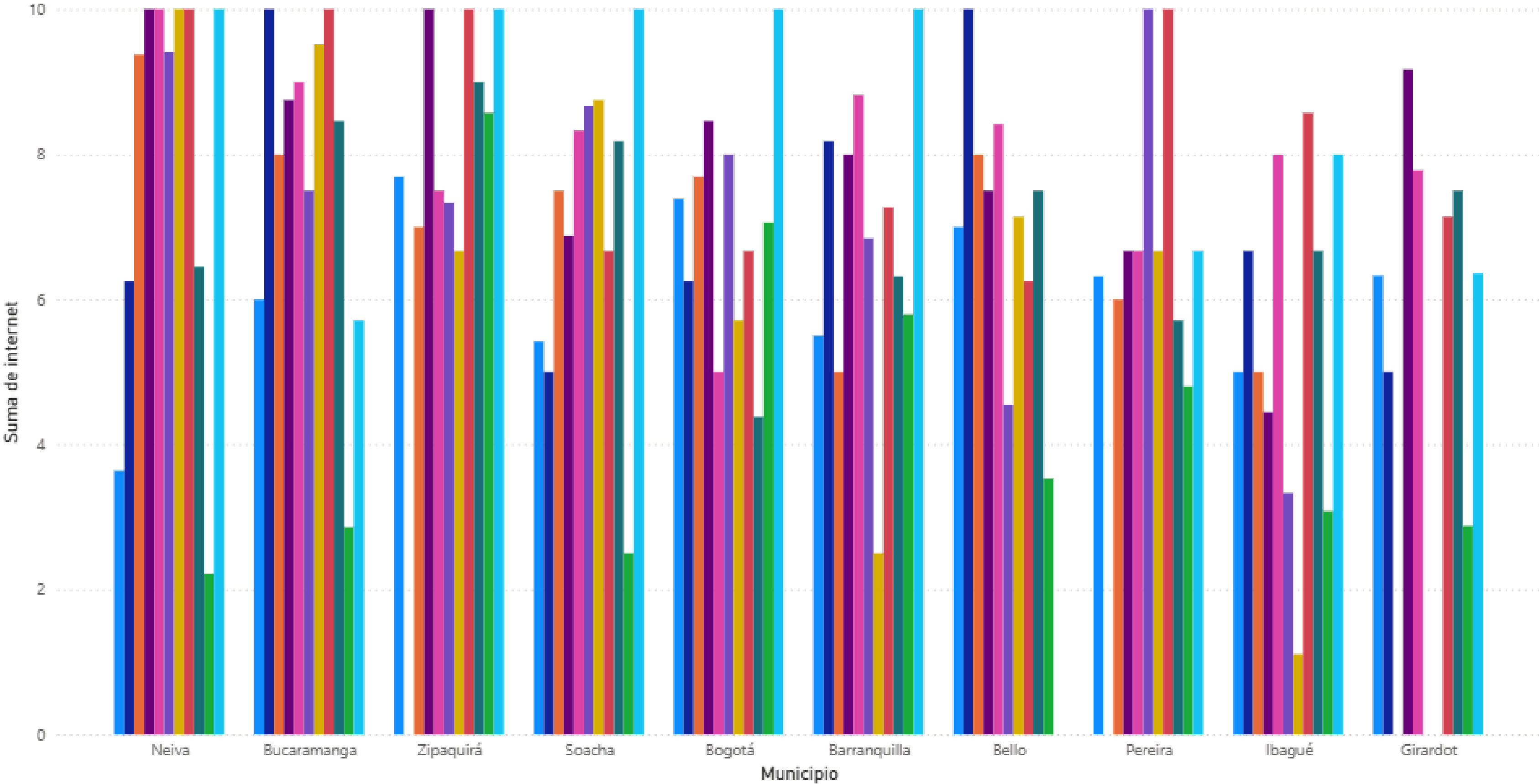
¿Cuáles sectores comerciales son los que más usan internet por ciudad?

En la mayoría de municipios los sectores que más usan internet son servicios, ropa, alimentos preparados y salud, lo que refleja la importancia de la digitalización en actividades vinculadas al consumo cotidiano y la atención al cliente. Por ejemplo, en Bogotá, Barranquilla y Pereira se destacan los servicios y la venta de ropa como los de mayor uso, mientras que en municipios como Soacha y Bucaramanga también sobresale el comercio de comida preparada. Esto muestra que los sectores con mayor interacción con consumidores y mayor necesidad de promoción o contacto digital son los que más dependen del internet, independientemente del tamaño de la ciudad.



Suma de internet por Municipio y Actividad

Actividad ComidaPreparada ferreteria Otras Papeleria Peluqueria productosinventario Ropa salud servicios Tienda vidanocturna



Relación entre población, actividad económica y uso de internet

El gráfico muestra que la relación entre población y uso de internet no es directamente proporcional, ya que mientras Bogotá concentra la mayor población, su nivel de conectividad no es tan alto frente a municipios con menor número de habitantes pero con mayor actividad económica o académica, como Pereira o Bucaramanga. Esto evidencia que el acceso digital no depende solo del tamaño poblacional, sino también del desarrollo económico y la infraestructura de cada territorio, lo que refleja una brecha digital entre municipios.



Promedio de poblacion y Suma de internet por Municipio

Promedio de poblacion Suma de internet

