

ACTIVIDADES

Sebastián Niklitschek

ACTIVIDAD I

- Ejecute la instrucción:
- `ggplot(data = mpg)`, ¿qué observa?
- ¿Cuántas columnas tiene el arreglo `mpg`? ¿Cuántas filas?
- ¿A qué hace referencia la variable `dvr` en el arreglo `mpg`?
- Cree un gráfico de dispersión para `hwy` vs `cyl`.
- ¿Qué ocurre si realiza un gráfico de dispersión para las variables `class` vs `drv`? ¿Por qué este tipo de visualización no resulta pertinente?

ACTIVIDAD II

- ¿Qué problema tiene el siguiente código?

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, color = "blue"))
```

- ¿Qué variables en el arreglo mpg son de tipo categórica? ¿Qué variables son continuas?
- Mapee variables continuas a los niveles color, tamaño y forma. ¿Observa alguna diferencia entre variables categóricas y continuas?
- ¿Qué ocurre si mapea la misma variable a diferentes niveles estéticos?
- ¿Qué hace el nivel estético “stroke”? ¿Cuándo es pertinente su uso?
- ¿Qué ocurre cuando asigna a un nivel algo diferente a una variable en el arreglo, por ejemplo aes(color = displ < 5)?

ACTIVIDAD III

- ¿Qué ocurre si usamos una variable continua en la función `facet_...`?
- ¿Qué significan las celdas vacías cuando usa `facet_grid(drv~cyl)`?
- Comente las diferencias entre los siguientes códigos. En particular explique el rol del “.”.

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_grid(drv ~ .)  
  
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_grid(. ~ cyl)
```

- ¿Qué hacen los argumentos `nrow` y `ncol`?

ACTIVIDAD IV

- Dibuje un gráfico de caja, un histograma y un “line chart”.
- Ejecute este código en su cabeza y haga una predicción de como se verá la salida:

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +  
  geom_point() +  
  geom_smooth(se = FALSE)
```

- ¿Qué hace el argumento “se” en la función geom_smooth()?

ACTIVIDAD V

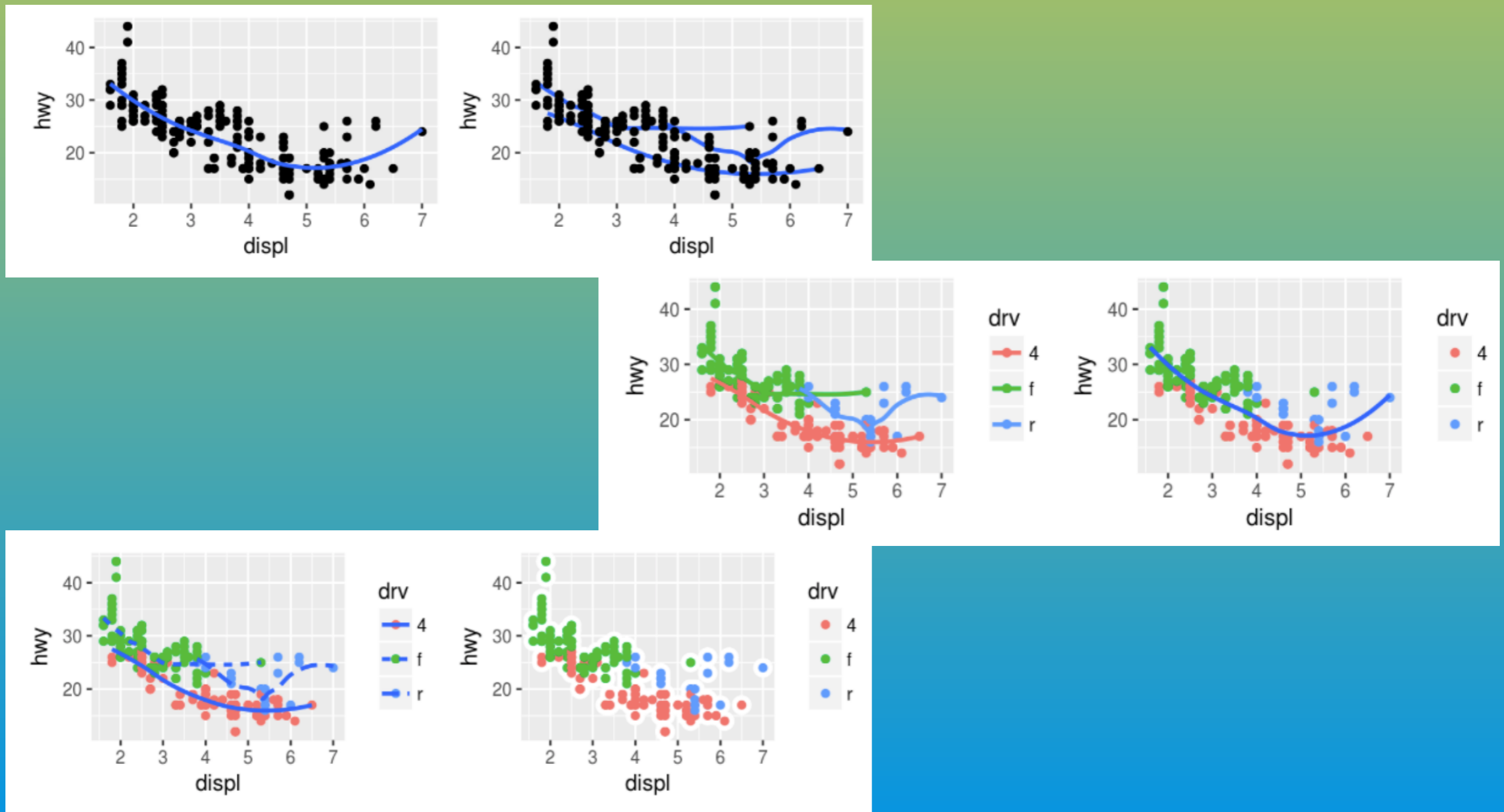
- Decida si las siguientes instrucciones generan o no el mismo gráfico:

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +  
  geom_point() +  
  geom_smooth()
```

```
ggplot() +  
  geom_point(data = mpg, mapping = aes(x = displ, y = hwy)) +  
  geom_smooth(data = mpg, mapping = aes(x = displ, y = hwy))
```

ACTIVIDAD VI

- Escriba el código en R necesario para generar las siguientes gráficas:



ACTIVIDAD VII

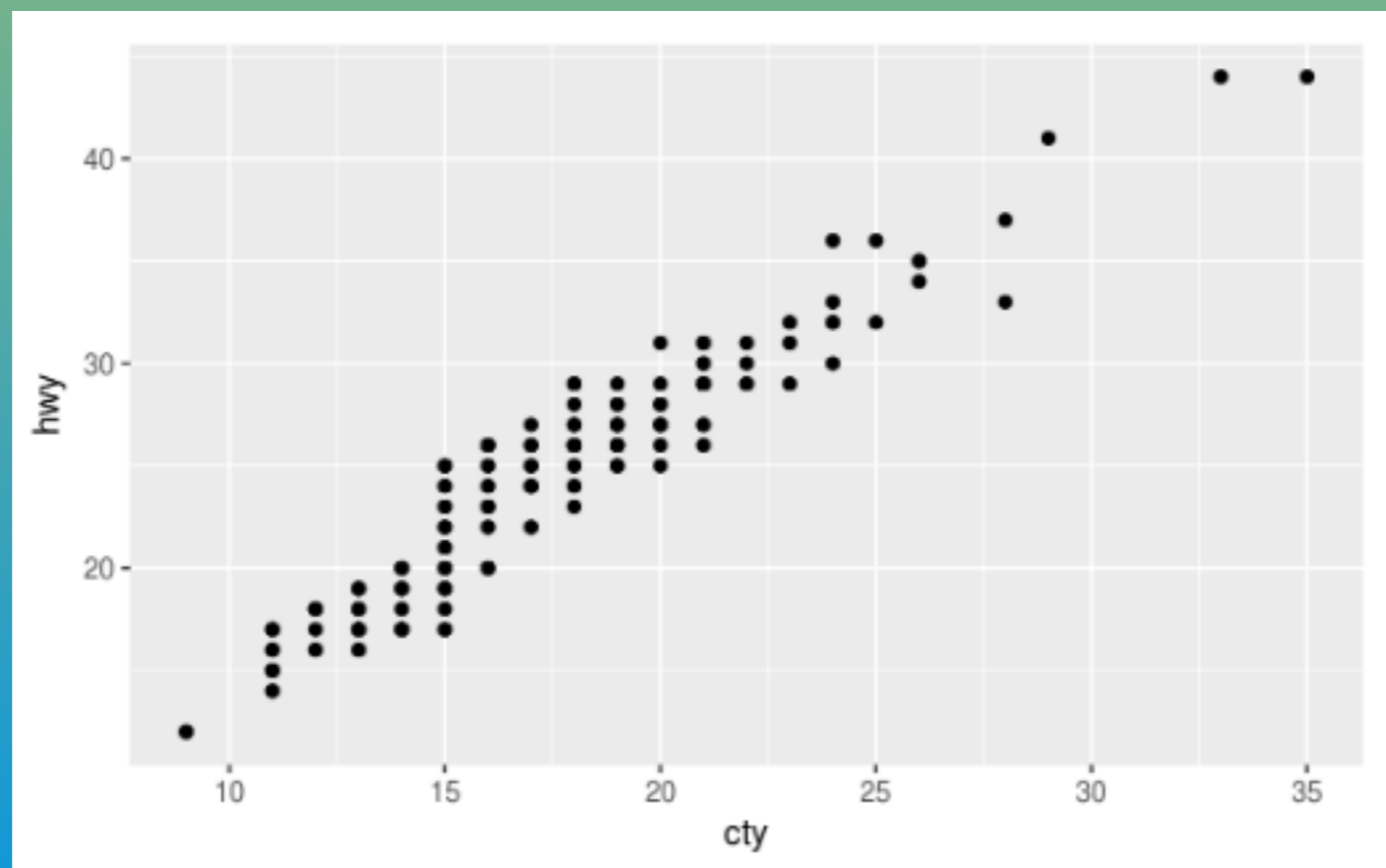
- ¿Cuál es la función geom asociada por defecto a `stat_summary()`? Dibuje el último gráfico utilizando una función de tipo geom.
- ¿Qué hace la función `geom_col()`? ¿Es diferente de `geom_bar()`?
- Genere una tabla con que de cuenta de los pares función geom-stat.
- ¿Qué variables calcula la función `geom_smooth()`?
- ¿Cuál es el problema con los siguientes gráficos?

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, y = ..prop..))  
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = color, y = ..prop..))
```


ACTIVIDAD VIII

- ¿Cuál es el problema con el siguiente gráfico y como puede solucionarlo?

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +  
  geom_point()
```



ACTIVIDAD IX

- ¿Qué parámetro de `geom_jitter()` controla el nivel de ruido?
- Compare y contraste `geom_jitter()` con `geom_count()`.
- ¿Cuál es el argumento por defecto para la estética “position”?
Construya una visualización de `mpg` que de cuenta de ello.

ACTIVIDAD X

- Convierta un gráfico de barras divididas en un gráfico de torta utilizando `coord_polar()`; (Siempre me he preguntado por qué no se llaman gráficos de pizza)
- ¿Qué hace la función `labs()`?
- ¿Cuál es la diferencia entre las funciones `coord_quickmap()` y `coord_map`?
- ¿Qué puede enseñarnos el siguiente gráfico acerca de la relación entre `hwy` y `city`? ¿Por qué es importante utilizar la función `coord_fixed()`? ¿Qué función cumple `geom_abline()`?

```
ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +  
  geom_point() +  
  geom_abline() +  
  coord_fixed()
```

DESAFÍO

- ▶ La librería gapminder contiene un extracto del conjunto de datos GapMinder;
- ▶ Para cada uno de 142 países, contiene los valores de expectativa de vida, PIB per cápita y población, medidos cada cinco años desde 1952 hasta el 2007;
- ▶ El desafío consiste en responder a las siguientes preguntas:
 - ¿Cómo se relacionan el PIB per cápita y la expectativa de vida?
 - ¿Ha cambiado esta relación con los años?
 - ¿Es diferente esta relación para cada continente?

VISUALIZACIÓN DE DATOS CON GGLOT2

Hoja de referencia

- <https://www.rstudio.com/wp-content/uploads/2015/04/ggplot2-spanish.pdf>