

Tarea y Solución

Pregunta 1

El siguiente gráfico que genera el código de R es correcto, pero puede mejorarse. ¿Qué cosas añadirías para mejorarlo?

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

Solución

El gráfico es muy pobre, tiene los puntos uno encima de otros (estaría mejor agitado con jitter) y seguro que con una capa `geom_smooth` veríamos una tendencia en los mismos. Más adelante verás más cosas acerca de este gráfico y acabarás sacando una conclusión por ti mismo.

Pregunta 2

Investiga la documentación de `geom_jitter()`. ¿Qué parámetros controlan la cantidad de ruido aleatorio (jitter)?

Solución

La agitación (jitter) se controla por:

- **width:** Amount of vertical and horizontal jitter. The jitter is added in both positive and negative directions, so the total spread is twice the value specified here. If omitted, defaults to 40% of the resolution of the data: this means the jitter values will occupy 80% of the implied bins. Categorical data is aligned on the integers, so a width or height of 0.5 will spread the data so it's not possible to see the distinction between the categories.
- **height:** Amount of vertical and horizontal jitter. The jitter is added in both positive and negative directions, so the total spread is twice the value specified here. If omitted, defaults to 40% of the resolution of the data: this means the jitter values will occupy 80% of the implied bins. Categorical data is aligned on the integers, so a width or height of 0.5 will spread the data so it's not possible to see the distinction between the categories.

Pregunta 3

Compara las funciones `geom_jitter` contra `geom_count` y busca semejanzas y diferencias entre ambas.

Solución

- `geom_point` se puede usar para hacer puntos más grandes de forma proporcional al número de observaciones de cada uno.
- `geom_jitter` agita los datos para que todos aparezcan visibles en lugar de alineados en una cuadrícula.

Pregunta 4

¿Cuál es el valor por defecto del parámetro `position` de un `geom_boxplot`? Usa el dataset de `diamonds` o de `mpg` para hacer una visualización que lo demuestre.

Solución

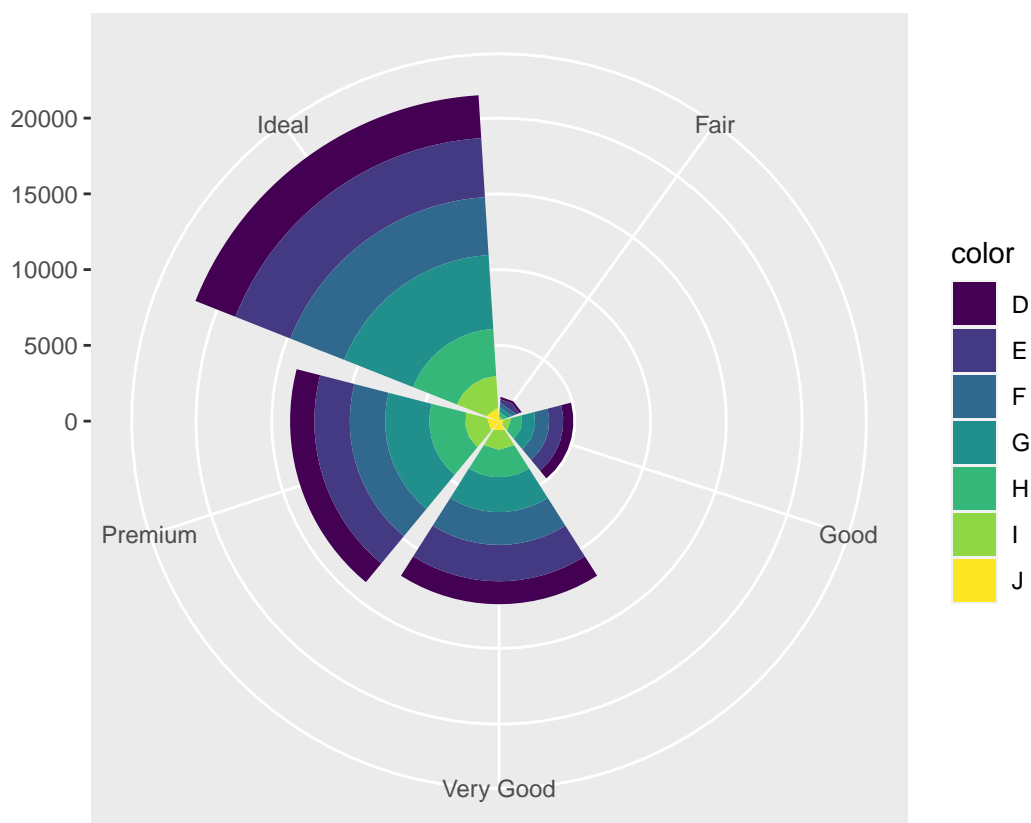
El valor por defecto, según la documentación de R, es `position = "dodge"`

Pregunta 5

Convierte un diagrama de barras apilado en un diagrama de sectores o de tarta usando la función `coord_polar()`.

Solución

```
ggplot(data = diamonds) +  
  geom_bar(mapping = aes(x = cut, fill = color)) +  
  theme(aspect.ratio = 1) +  
  labs(x = NULL, y = NULL) +  
  coord_polar()
```



Pregunta 6

¿Qué hace la función `labs()`? Lee la documentación y explícalo correctamente.

Solución

Sirven para modificar el título (title, caption y subtitle), el nombre de los ejes (x e y) y la leyenda del plot.

Good labels are critical for making your plots accessible to a wider audience. Ensure the axis and legend labels display the full variable name. Use the plot `title` and `subtitle` to explain the main findings. It's common to use the `caption` to provide information about the data source.

Pregunta 7

¿En qué se diferencian las funciones `coord_quickmap()` y `coord_map()`?

Solución

- `coord_map` projects a portion of the earth, which is approximately spherical, onto a flat 2D plane using any projection defined by the `mapproj` package. Map projections do not, in general, preserve straight lines, so this requires considerable computation.
- `coord_quickmap` is a quick approximation that does preserve straight lines. It works best for smaller areas closer to the equator.

Pregunta 8

Investiga las coordenadas `coord_fixed()` e indica su función.

Solución

Son coordenadas cartesianas con aspect ratio fijo.

A fixed scale coordinate system forces a specified ratio between the physical representation of data units on the axes. The ratio represents the number of units on the y-axis equivalent to one unit on the x-axis. The default, `ratio = 1`, ensures that one unit on the x-axis is the same length as one unit on the y-axis. Ratios higher than one make units on the y axis longer than units on the x-axis, and vice versa. This is similar to `eqsplot`, but it works for all types of graphics.

Pregunta 9

Investiga la geometría de la función `geom_abline()`, `geom_vline()` y `geom_hline()` e indica su función respectivamente.

Solución

Sirve para representar rectas en el plano:

These geoms add reference lines (sometimes called rules) to a plot, either horizontal, vertical, or diagonal (specified by slope and intercept). These are useful for annotating plots.

Pregunta 10

¿Qué nos indica el gráfico siguiente acerca de la relación entre el consumo en ciudad y en autopista del dataset de mpg?

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

Solución

La relación del consumo en ciudad y autopista es lineal: si un coche consume más por autopista, es normal que también consuma más por ciudad. El aspect ratio fixed nos ayuda en la representación para no confundirnos y pensar que una variable crece más rápida que la otra en el mismo. Además, si nos fijamos en el gráfico, vemos que los puntos están por encima de la recta $y=x$, por tanto concluimos que el número de millas que podemos hacer en autopista con un galón de combustible es siempre superior al número de millas que haremos con ese mismo galón en ciudad (es decir, consumimos menos en autopista que en ciudad)