

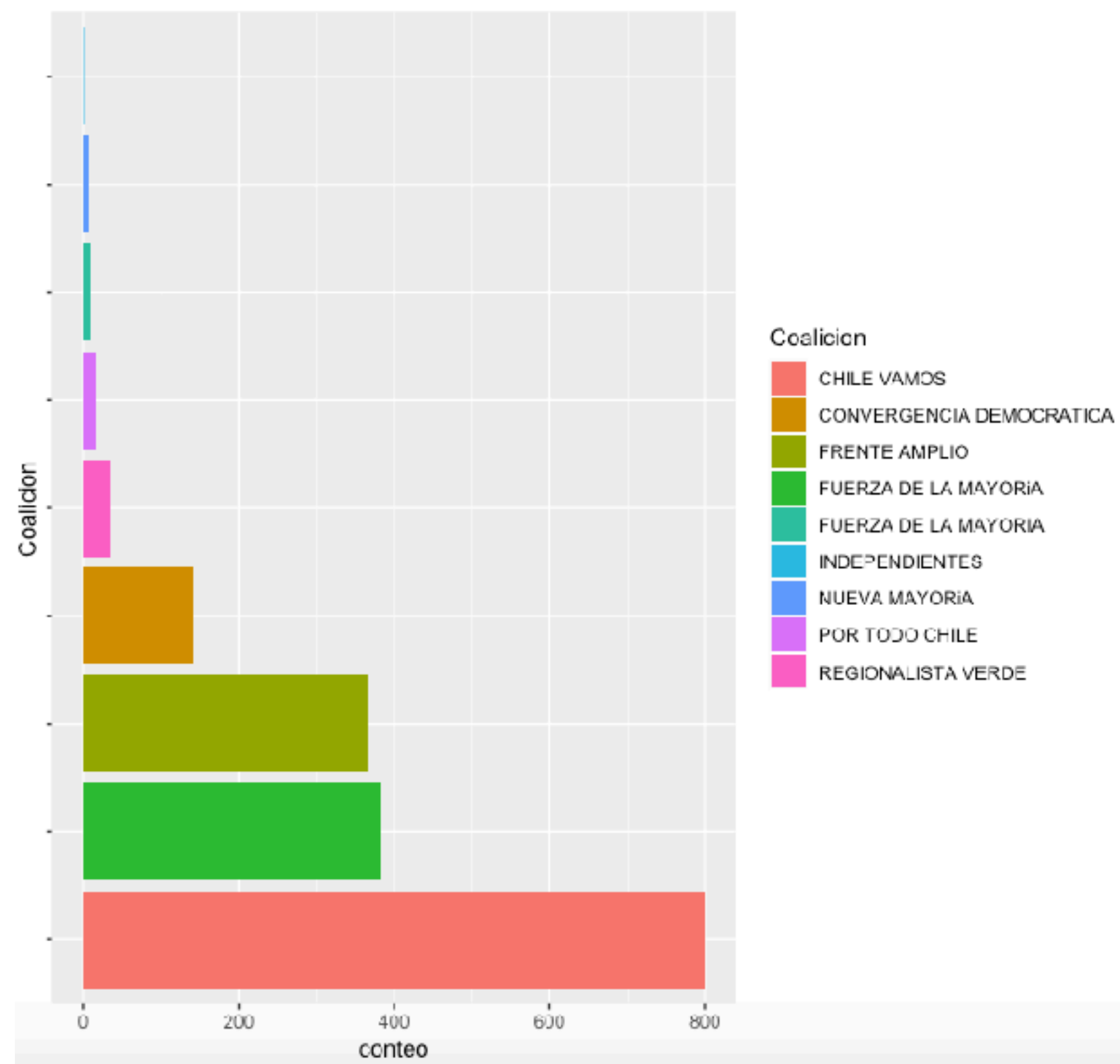
SEBASTIAN NIKLITSCHKE

GRAMÁTICA GRÁFICA

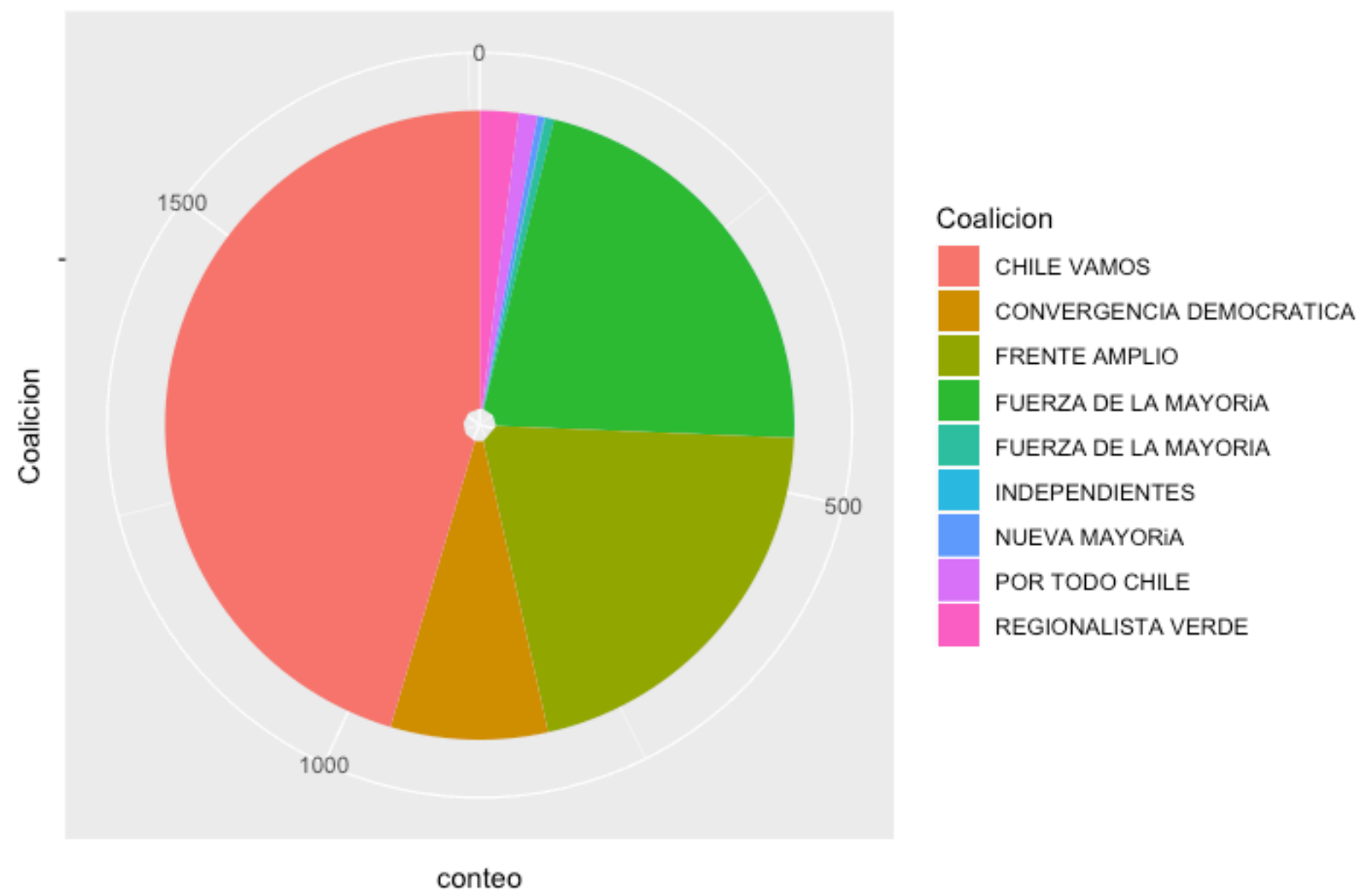
ABSTRACCIÓN

- ▶ Diferentes tipos de gráficos pueden, a primera vista, parecer completamente distintos. Pero, de hecho, los gráficos comparten muchos elementos comunes, como los sistemas de coordenadas o el uso de formas geométricas para representar datos.
- ▶ Al realizar diferentes elecciones visuales (coordenadas cartesianas o polares, puntos o líneas o barras para representar datos), podemos resaltar diferentes aspectos de los mismos datos.

UN PRIMER EJEMPLO



UN SEGUNDO EJEMPLO



¿DIFERENCIAS?

- ▶ El gráfico circular enfoca al lector en grandes porcentajes y lo alienta a pensar en el total de audiencias como una cantidad finita que se está distribuyendo a diferentes grupos. El gráfico de barras apiladas proporciona la misma información, pero facilita la medición precisa del tamaño de cada categoría.

¿DIFERENCIAS?

- ▶ A menudo hablamos de tipos de gráficos (gráficos de barras, gráficos circulares, gráficos de dispersión) como si no estuvieran relacionados, pero la mayoría de los gráficos comparten muchos aspectos de su estructura.
- ▶ Podemos pensar en los gráficos como representaciones visuales de datos, junto con etiquetas que aclaran el significado.
- ▶ Al igual que la gramática de un lenguaje le permite combinar palabras en oraciones significativas, una gramática de gráficos proporciona una estructura para combinar elementos gráficos en figuras que muestran datos de manera significativa.

GRAMÁTICA DE LOS GRAFOS

- ▶ La gramática de los gráficos fue introducida originalmente por Leland Wilkinson a fines de la década de 1990, y fue popularizada por Hadley Wickham con ggplot, una biblioteca gráfica en R basada en la gramática de los gráficos.
- ▶ La generación de visualizaciones con ggplot puede ser poco natural en un comienzo, pero se facilita mucho si se comprende la teoría detrás de la gramática de los gráficos

¿GRAMÁTICA?

- ▶ En la gramática de un lenguaje, las palabras tienen diferentes partes del habla, que desempeñan diferentes roles en la oración. De forma análoga, la gramática de los gráficos separa un gráfico en diferentes capas. Estas son capas en un sentido literal: puede considerarlas como hojas de transparencia para un proyector, cada una con una parte del gráfico, que se puede organizar y combinar de varias maneras.

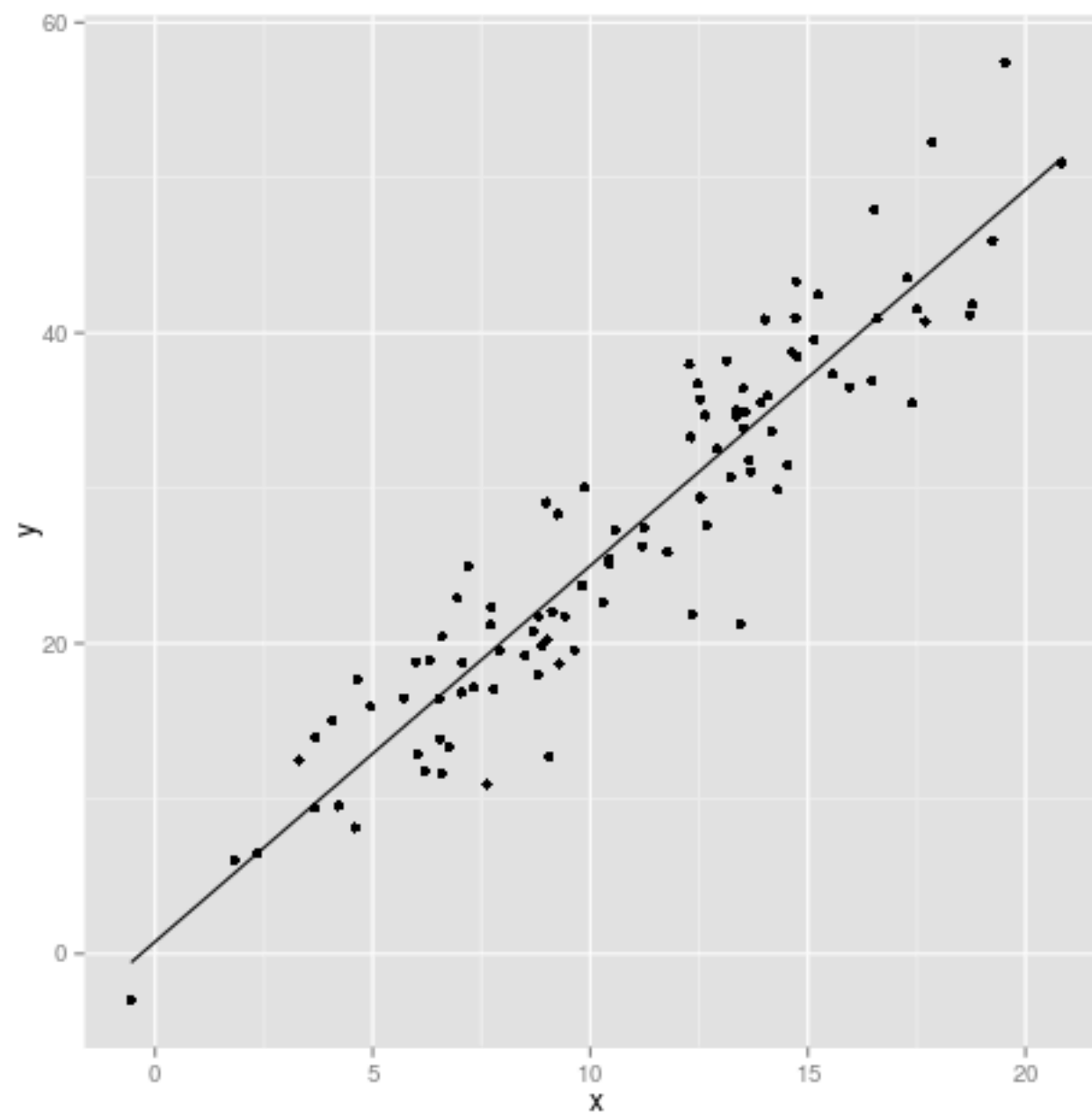
¿QUÉ HAY EN CADA CAPA?

- ▶ Pensemos en un ejemplo. Supongamos que tenemos un conjunto de datos con una variable independiente, x , y una variable dependiente, y .
- ▶ Si realizamos una regresión lineal simple, también podemos calcular los valores pronosticados para y en valores específicos de x (podemos llamar a estas predicciones \hat{y}). Usando estos datos, queremos hacer un diagrama de dispersión con una línea de mejor ajuste. ¿Cuáles son los elementos de esta trama?

TRAMA

- ▶ Nuestra trama está compuesta por:
 - ▶ Los datos en sí (x , y , y la mejor predicción de ajuste, y)
 - ▶ Puntos en el diagrama de dispersión que representan la relación entre x e y
 - ▶ La línea que representa la relación entre x e y (la línea de mejor ajuste)
 - ▶ La escala de los datos (lineales).
 - ▶ El sistema de coordenadas (cartesiano).

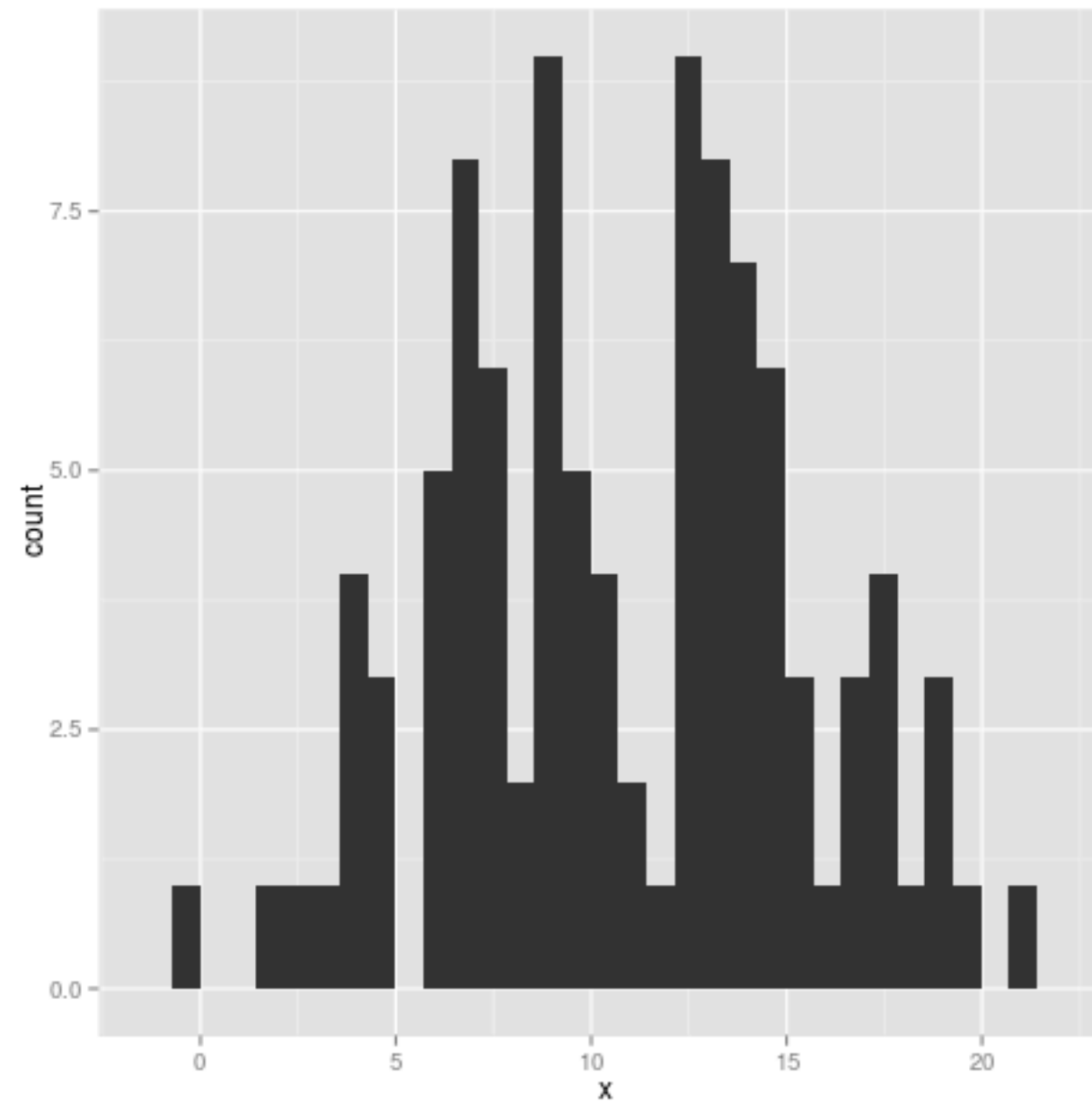
RESULTADO



TRAMA 2

- ▶ ¿Qué pasa si queremos hacer un histograma de la distribución de x ?
- ▶ Los elementos de nuestra trama son:
 - ▶ Los datos x ,
 - ▶ Barras que representan la frecuencia de x en diferentes valores de x ,
 - ▶ La escala de los datos (lineales),
 - ▶ El sistema de coordenadas (cartesiano).

RESULTADO



OBSERVACIÓN

- ▶ Claramente, hay muchos componentes similares en ambas tramas, esto, producto de que hay componentes similares entre las visualizaciones. Estos componentes similares son "partes del discurso", las piezas de un gráfico que podemos usar para contar una historia.
- ▶ Revisaremos cada una de modo de comprenderlas mejor.

ELEMENTOS DE LA TRAMA

DATOS

TABLAS DE DATOS

- ▶ Antes de hablar de gramática de los gráficos, es necesario conocer y entender el formato y la estructura de nuestros datos.
- ▶ La gramática se nutre a partir de tablas de datos, esto es importante porque no siempre los conjuntos de datos se encuentran ordenados de esta forma.
- ▶ data frame en R, pandas data frame en python.

TABLA

Sujeto.Pasivo	Fecha	Lobbista.representado	Lugar
schalper sepulveda diego	02-10-2018 10:00:00	asociación gremial nacional de organizaciones de feri...	Congreso, piso 12, oficina 1.
sepulveda soto alexis	01-10-2018 18:15:00	jorge hermann	Oficina comité Radical, Ex co
calisto aguila miguel angel	01-10-2018 15:30:00	uber b.v.	cercanía Ex Congreso Santia
calisto aguila miguel angel	01-10-2018 15:30:00	uber b.v.	cercanías Ex Congreso Santia
renteria moller rolando	01-10-2018 13:00:00	ong cruzada solidaria	ex congreso nacional Estació
renteria moller rolando	01-10-2018 12:30:00	asociación de directivos, profesionales, técnicos, adm...	ex congreso nacional Provid

IMPORTANTE

- ▶ Usualmente nuestra tabla de datos podría no contener las variables tal como nos gustaría utilizarlas para escribir nuestra historia.
- ▶ En ese caso, es importante el dominar las herramientas para manejo de datos. La librería dplyr en R y dfply en python son maravillosas para eso.

ELEMENTOS DE LA TRAMA

GEOMS

REPRESENTACIÓN VISUAL

- ▶ La parte más obvia de un gráfico es la representación visual de los datos en sí. A menudo, este es un objeto geométrico básico como un punto, una línea o una barra.
- ▶ En ggplot, cada uno de estos elementos se denomina "geom".
- ▶ Es posible múltiples piezas de información por capas de geoms (capa de dispersión + línea de mejor ajuste), o explorar los mismos datos visualizándolos con diferentes tipos de geoms.

UN EJEMPLO

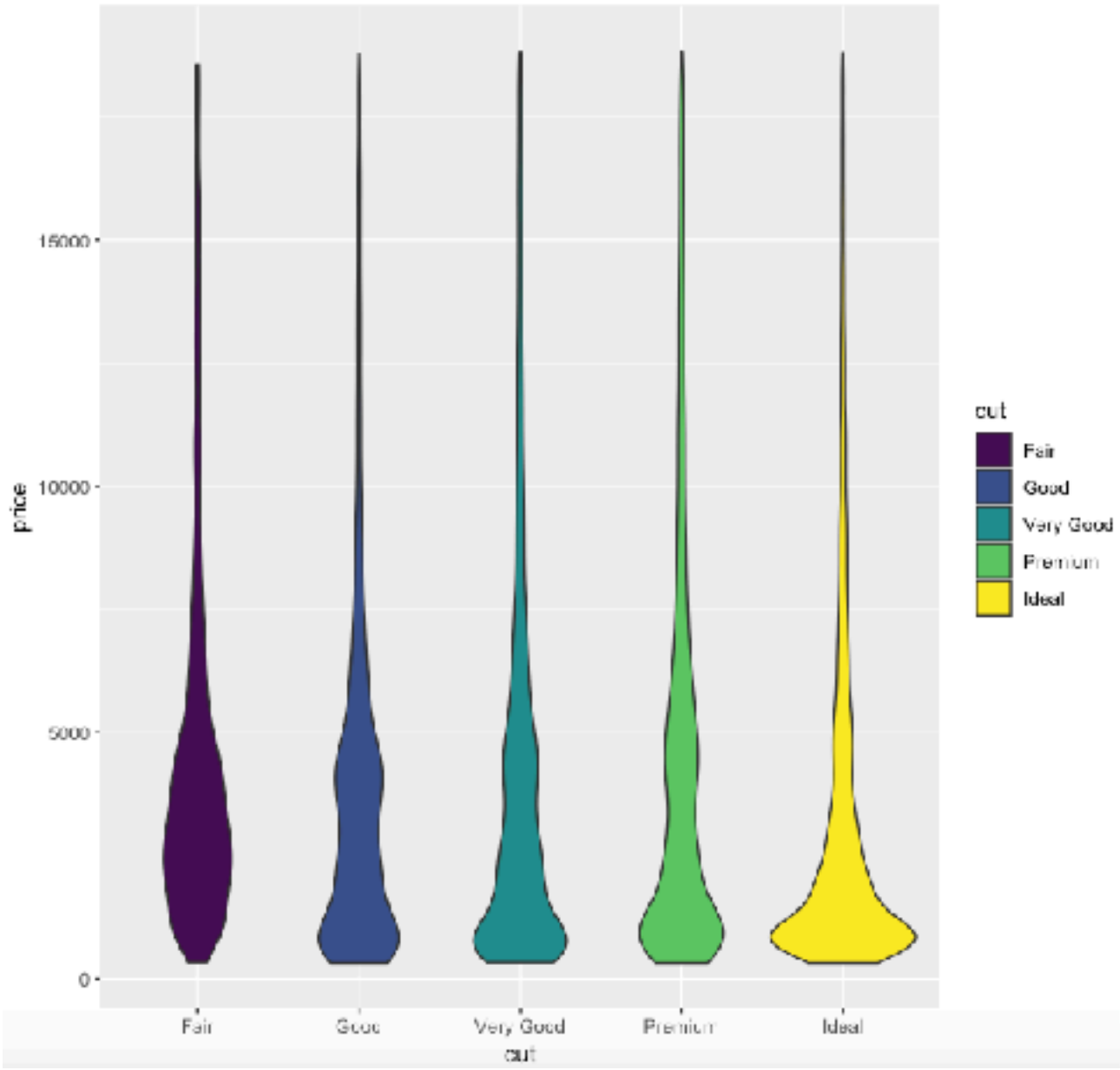
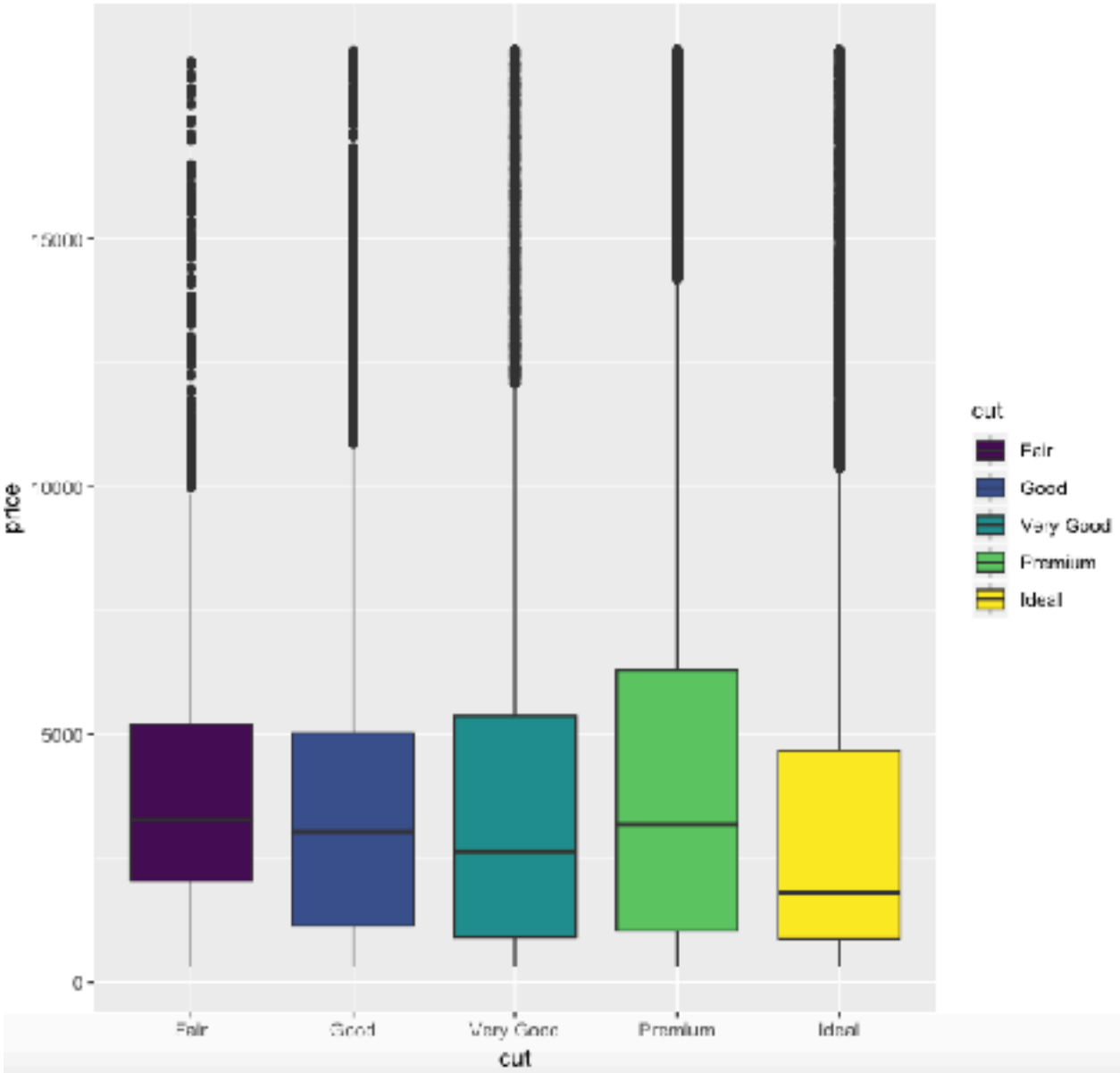
- ▶ Supongamos que tenemos un arreglo con ejemplos de diamantes, descritos por sus valores de quilates y precio.
- ▶ Podemos generar diferentes visualizaciones en función de qué objeto geom utilicemos.

```
# Primero generamos la gráfica
g <- ggplot(diamonds, aes(x = cut, y = price, fill = cut))

## gráficos de caja
gbox <- g + geom_boxplot()

# gráficos de violín permiten visualizar la distribución completa
gviol <- g + geom_violin()
```

DOS EJEMPLOS



ELEMENTOS DE LA TRAMA

ESCALA

ESCALAS – RE-ESCALAMIENTOS

- ▶ Muchas veces es necesario cambia la escala en la que se encuentran nuestros datos.
- ▶ Los humanos somos muy buenos para interpretar y comprar escalas lineales por lo que si la escala de nuestros datos es diferente, es una buena idea transformarla de modo que sea más fácilmente interpretable.

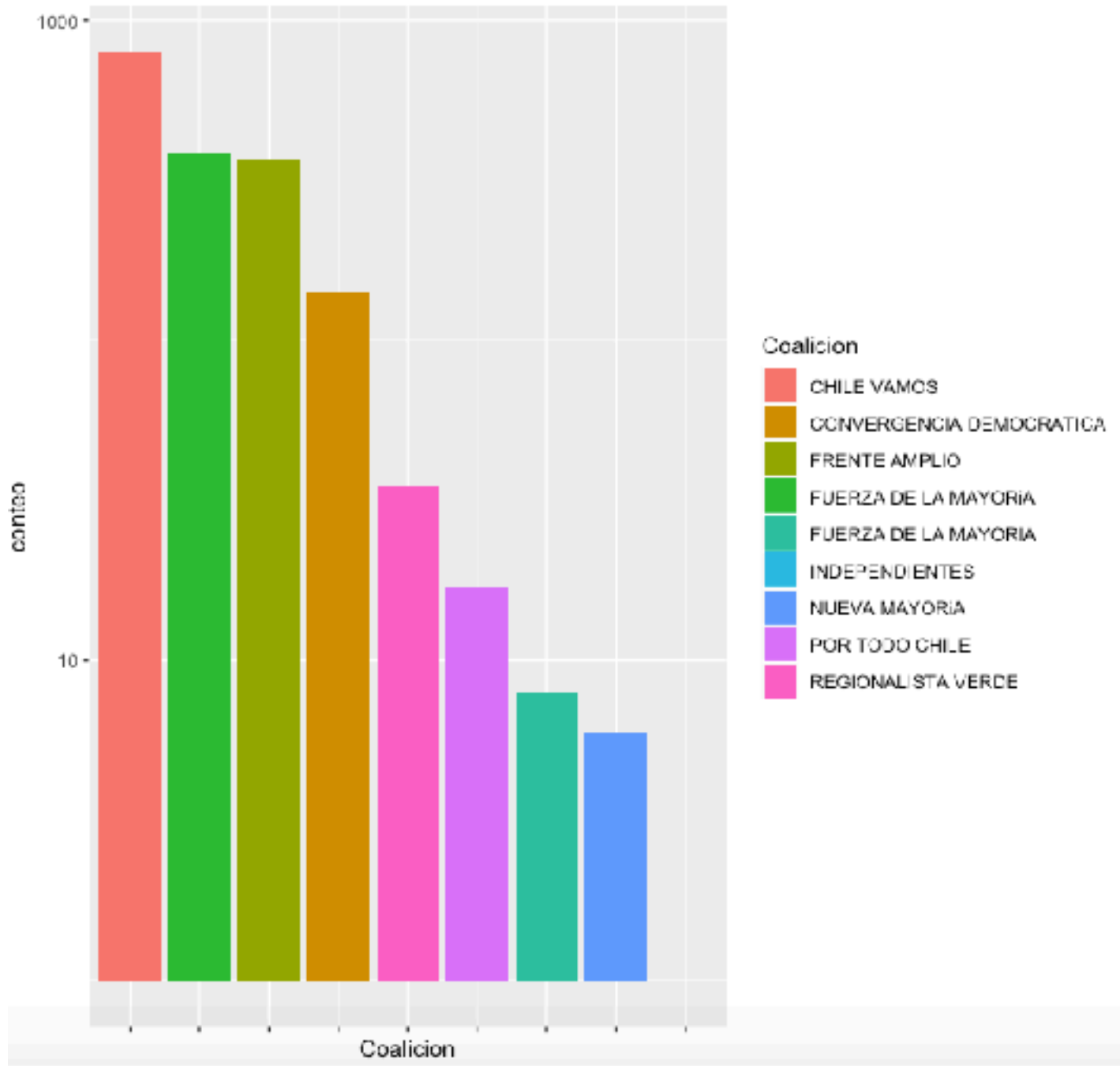
ELEMENTOS DE LA TRAMA

COORDENADAS

EJES COORDENADOS

- ▶ Estamos muy acostumbrados a pensar utilizando el sistema de coordenadas cartesiano, pero a veces las coordenadas polares tienen sentido.
- ▶ Los gráficos circulares son un uso común (aunque controvertido) de las coordenadas polares.
- ▶ También es posible cambiar las coordenadas de un gráfico de barras horizontales en lugar de uno vertical.

EJEMPLO



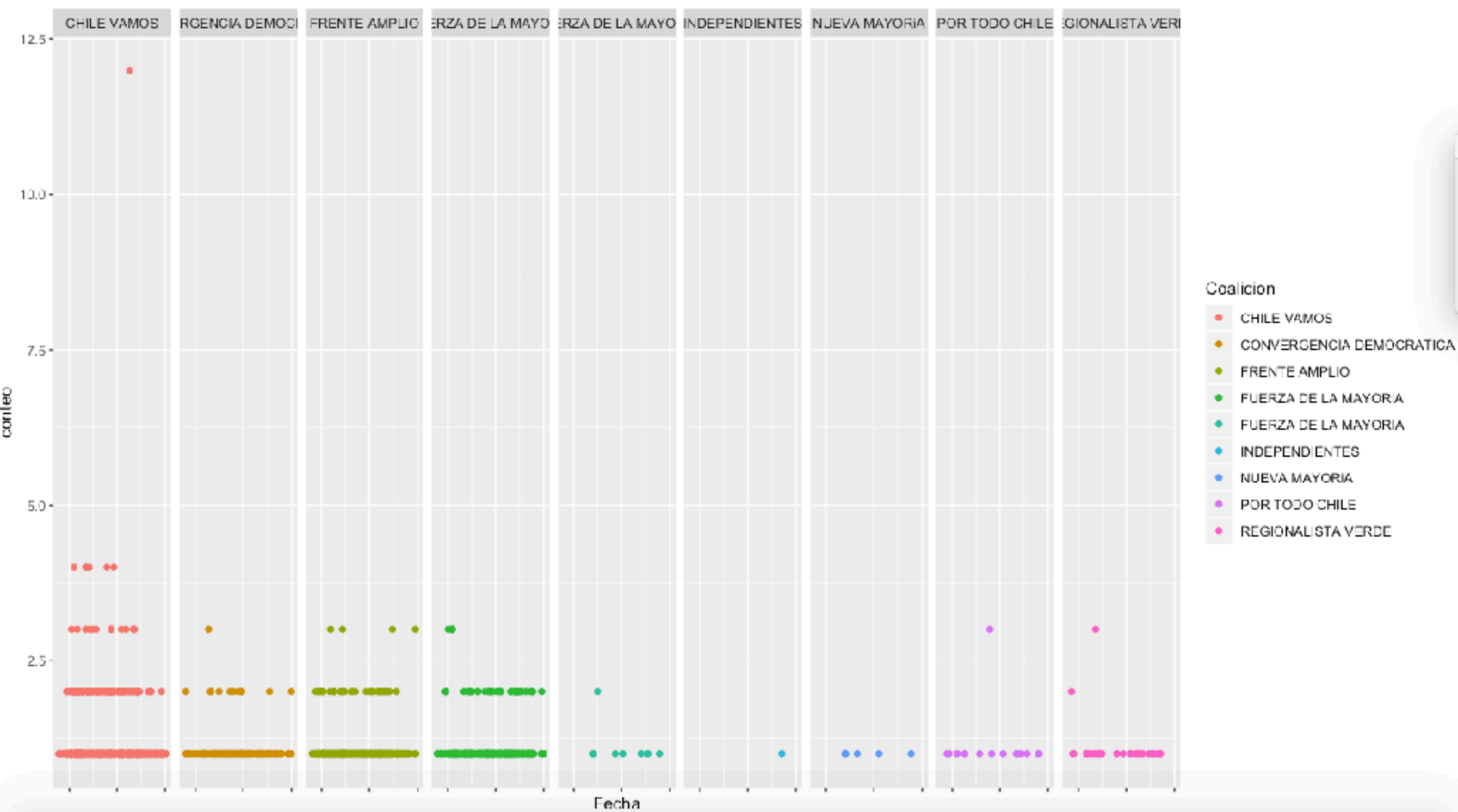
ELEMENTOS DE LA TRAMA

GRUPOS Y FACETAS

FACETAS

- ▶ Las facetas son una forma de dividir los datos en subparcelas en función de otro factor en los datos.
- ▶ Las facetas son una de las razones más convincentes para usar la gramática de los gráficos.

DEMODATA 2019



VENTAJAS DE LA GRAMÁTICA GRÁFICA

- ▶ Con una gramática con la que trabajar, podemos comunicar nuestras intenciones a ggplot de una manera más clara y concisa.
- ▶ Para antropomorfizar en gran medida, los gráficos de Base R grafican los datos, pero no "entienden" nada sobre cómo se estructuran los datos. Si desea dividir sus datos en grupos por otra variable, debe decirle específicamente qué hacer con cada fragmento de datos, es decir no entiende lo que significa "agrupado por otra variable".

VENTAJAS DE LA GRAMÁTICA GRÁFICA

- ▶ La gramática de los gráficos proporciona un lenguaje común entre la computadora y el usuario. ggplot entiende el concepto de un conjunto de datos y cómo se relacionan sus filas y columnas. Comprende que un gráfico circular y un gráfico de barras apiladas son el mismo gráfico con diferentes sistemas de coordenadas. Comprende la idea de los gráficos facetados como una sola unidad visual. Y lo más importante, hace que esas ideas y relaciones sean visibles para el usuario, para que sea sencillo cambiar entre diferentes elementos visuales para representar los datos.

CONCLUSIONES

- ▶ La comunicación es la idea central en el trabajo aquí.
- ▶ Cuando tienes un lenguaje estructurado para gráficos, es mucho más fácil pensar y hablar sobre ellos.
- ▶ Es un gran marco mental para cuando estás tratando de decidir cómo mostrar datos para ti mismo o para una presentación.
- ▶ Si no está seguro de cómo desea mostrar sus datos, proporciona una manera concisa y coherente de moverse entre las diferentes representaciones posibles.