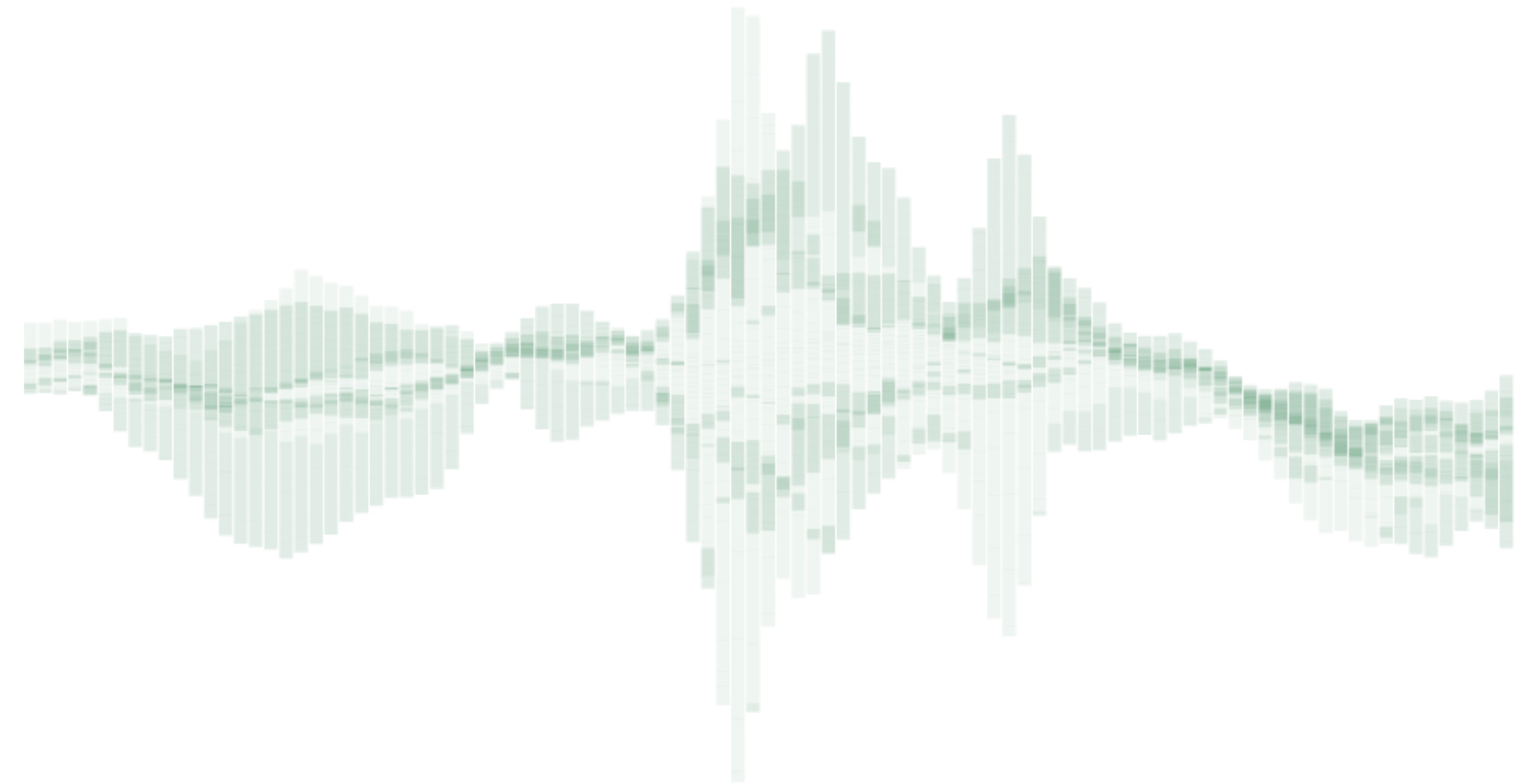




Distribuciones



25 de marzo de 2024

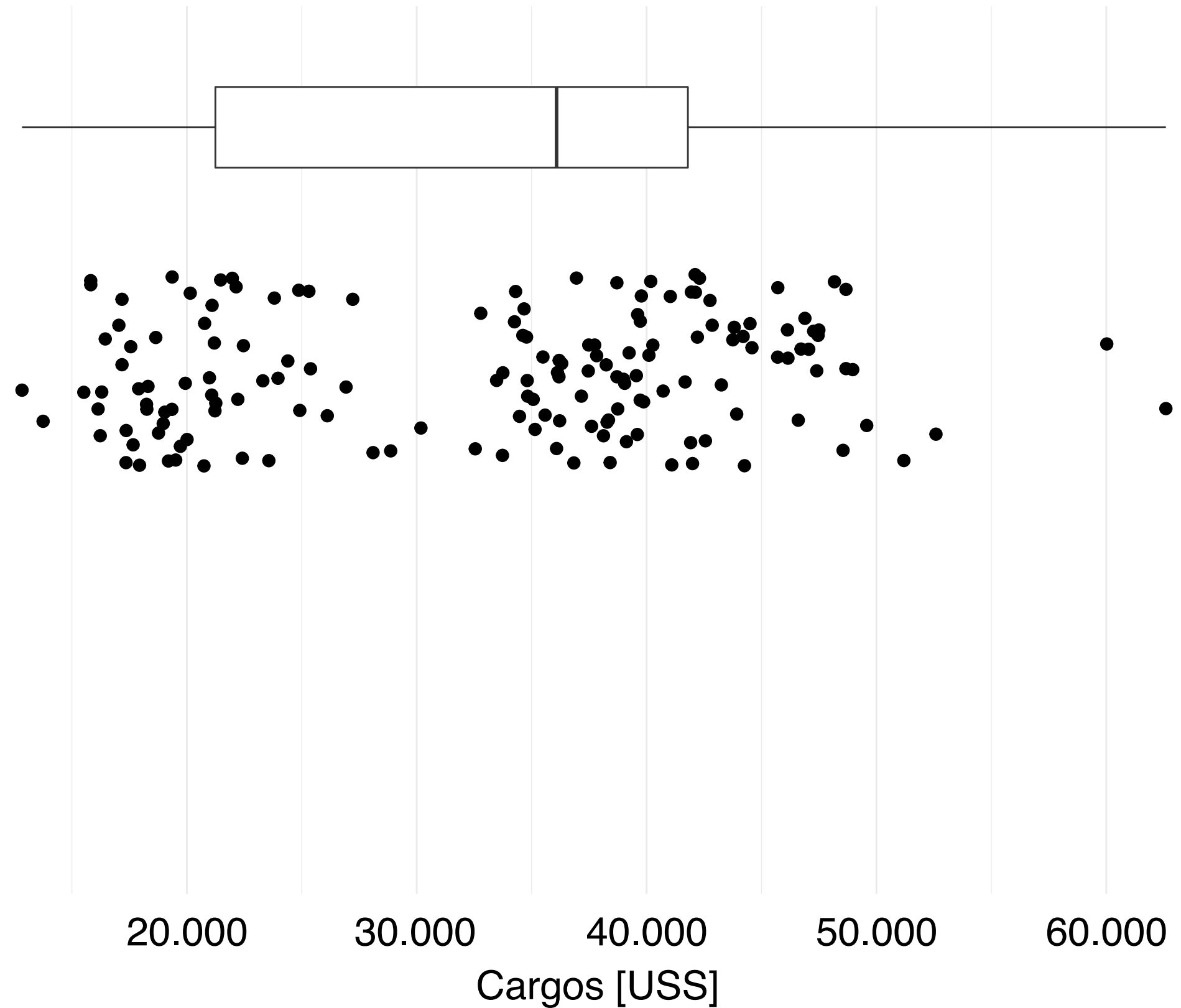
Hasta ahora

- 🌿 Exploración del dataset
- 🌿 Manipulación (orden, filtrado, agrupamiento)
- 🌿 Métricas de resumen + graficarlas (boxplot)

insurance

costo seguro médico
hombres, fumadores

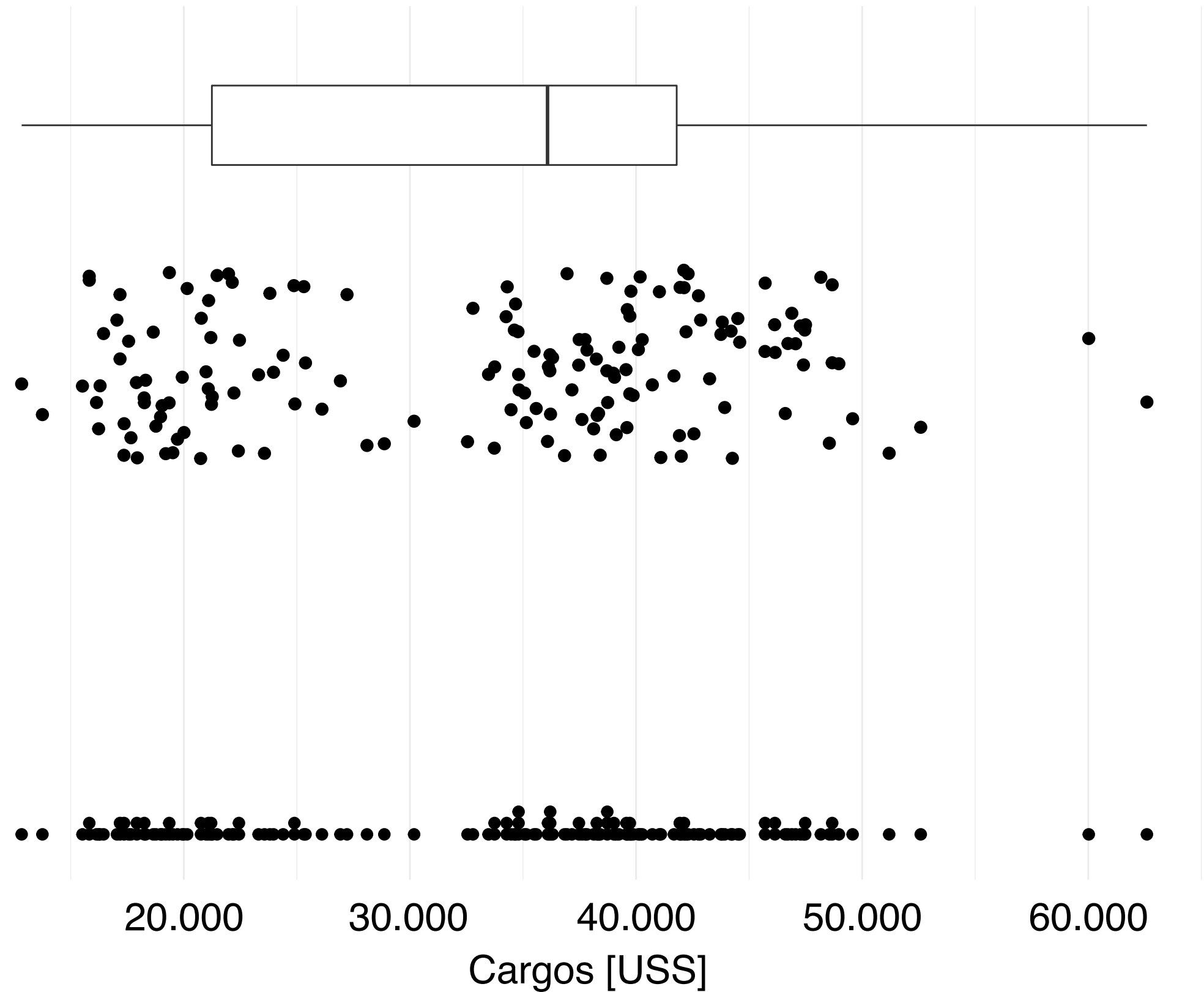
`geom_boxplot()` +
`geom_jitter()`



insurance

costo seguro médico
hombres, fumadores

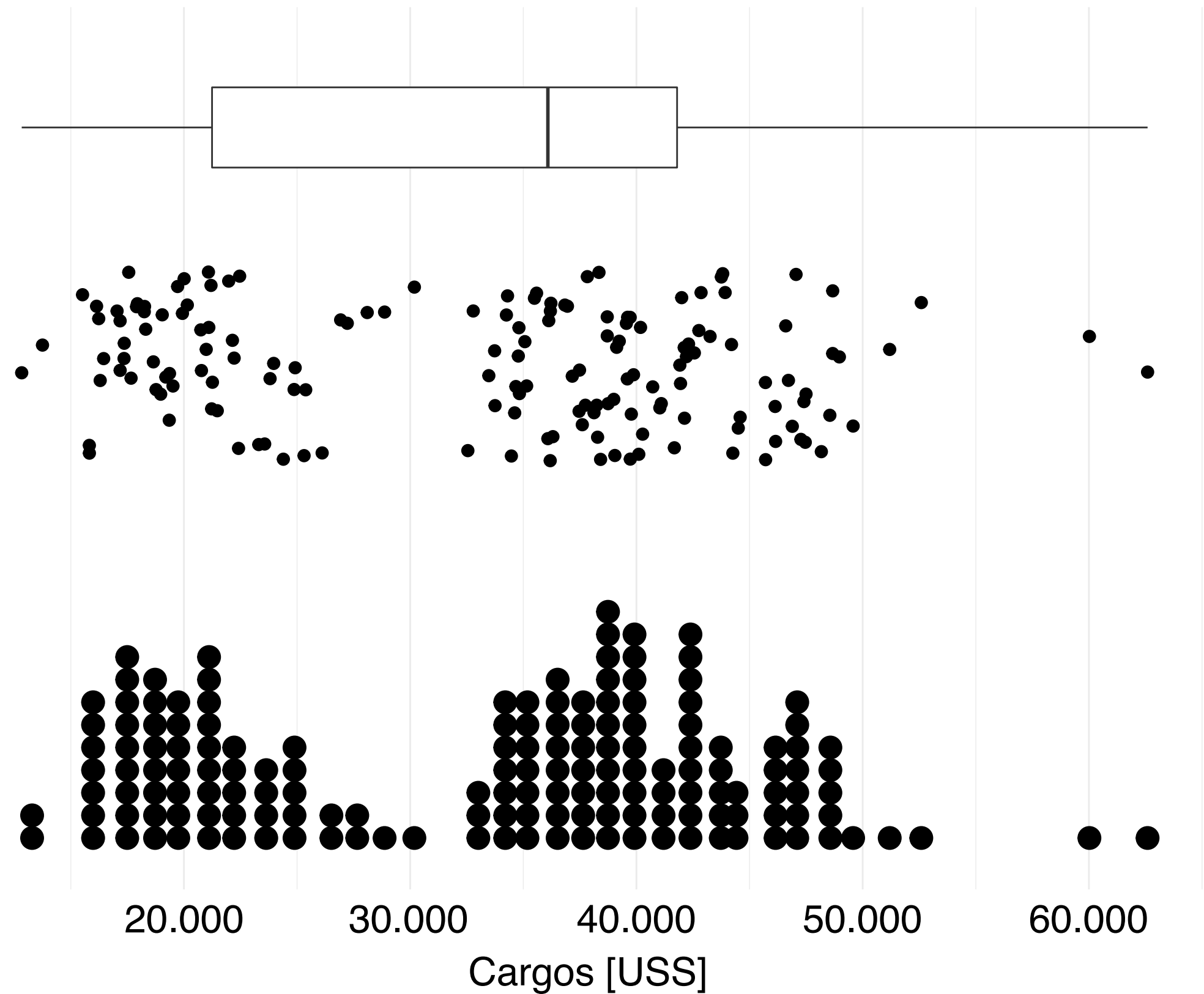
geom_boxplot() +
geom_jitter() +
geom_dotplot()



insurance

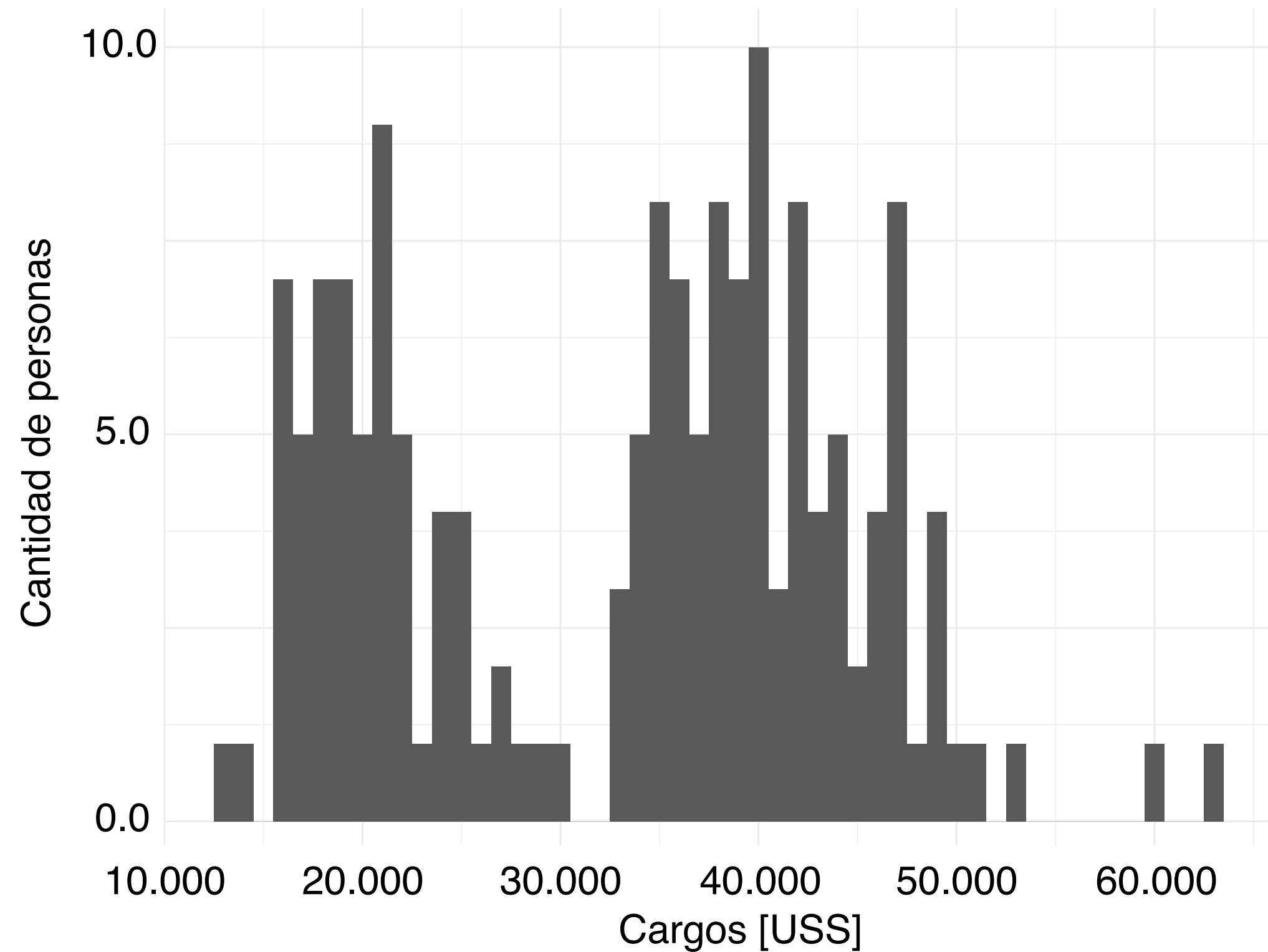
costo seguro médico
hombres, fumadores

geom_boxplot() +
geom_jitter() +
geom_dotplot()



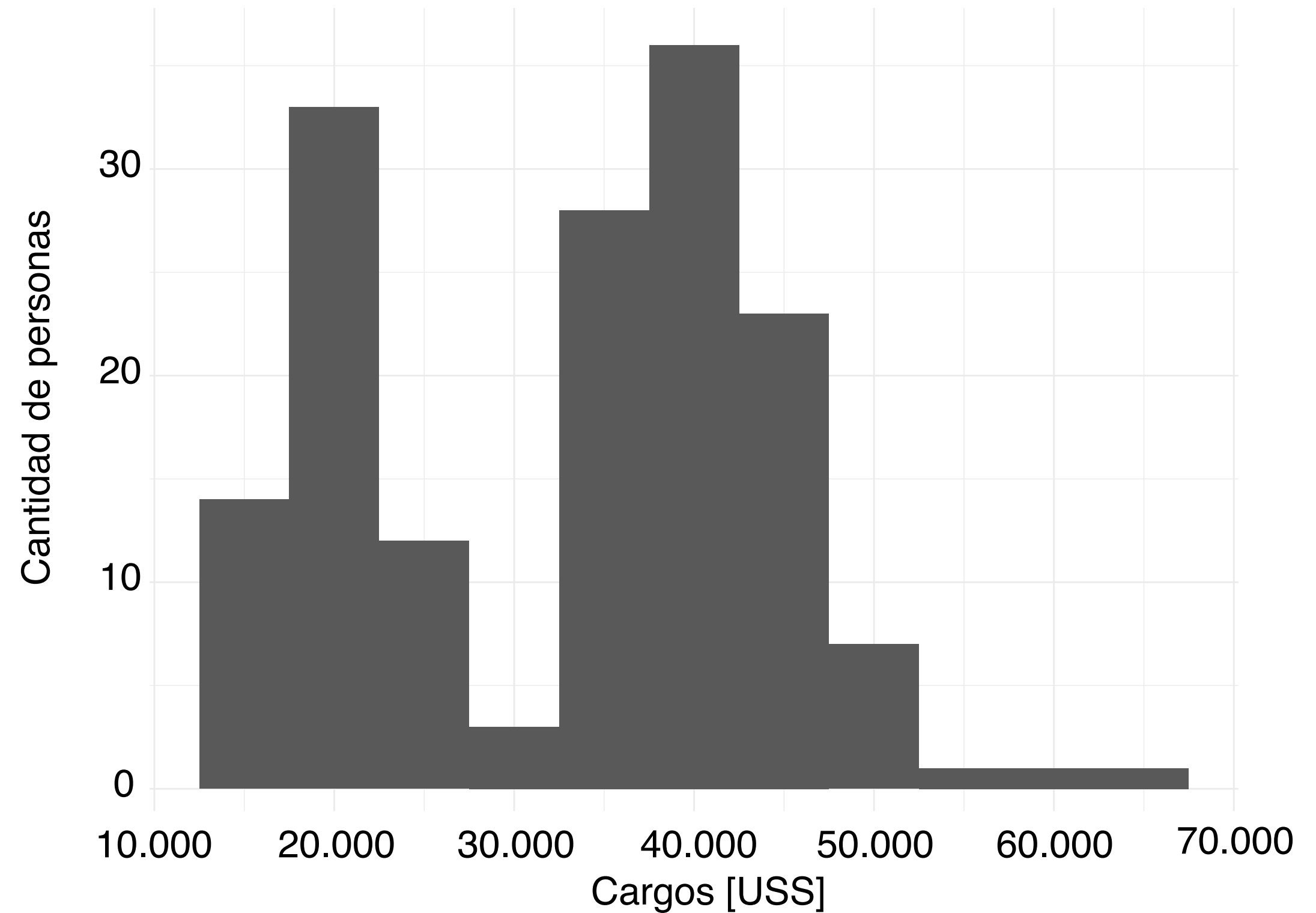
histograma

`geom_histogram(
 binwidth = 1000)`



histograma

```
geom_histogram(  
  binwidth = 5000)
```



comparar distintos tamaños de conjuntos

```
geom_histogram(  
  fill= smoker)
```

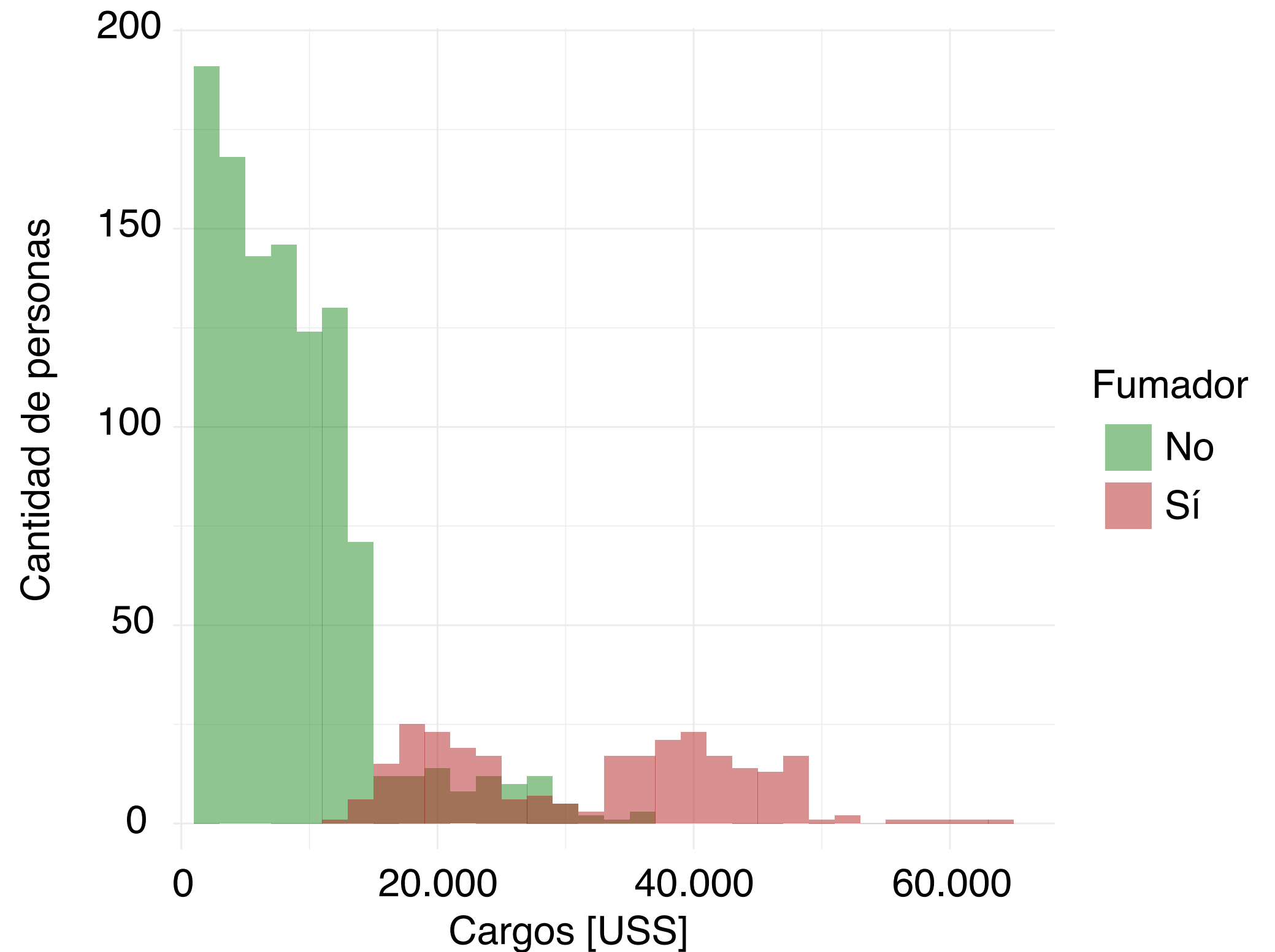


gráfico de densidad

```
geom_histogram(  
  fill= smoker,  
  y = ..density..)
```

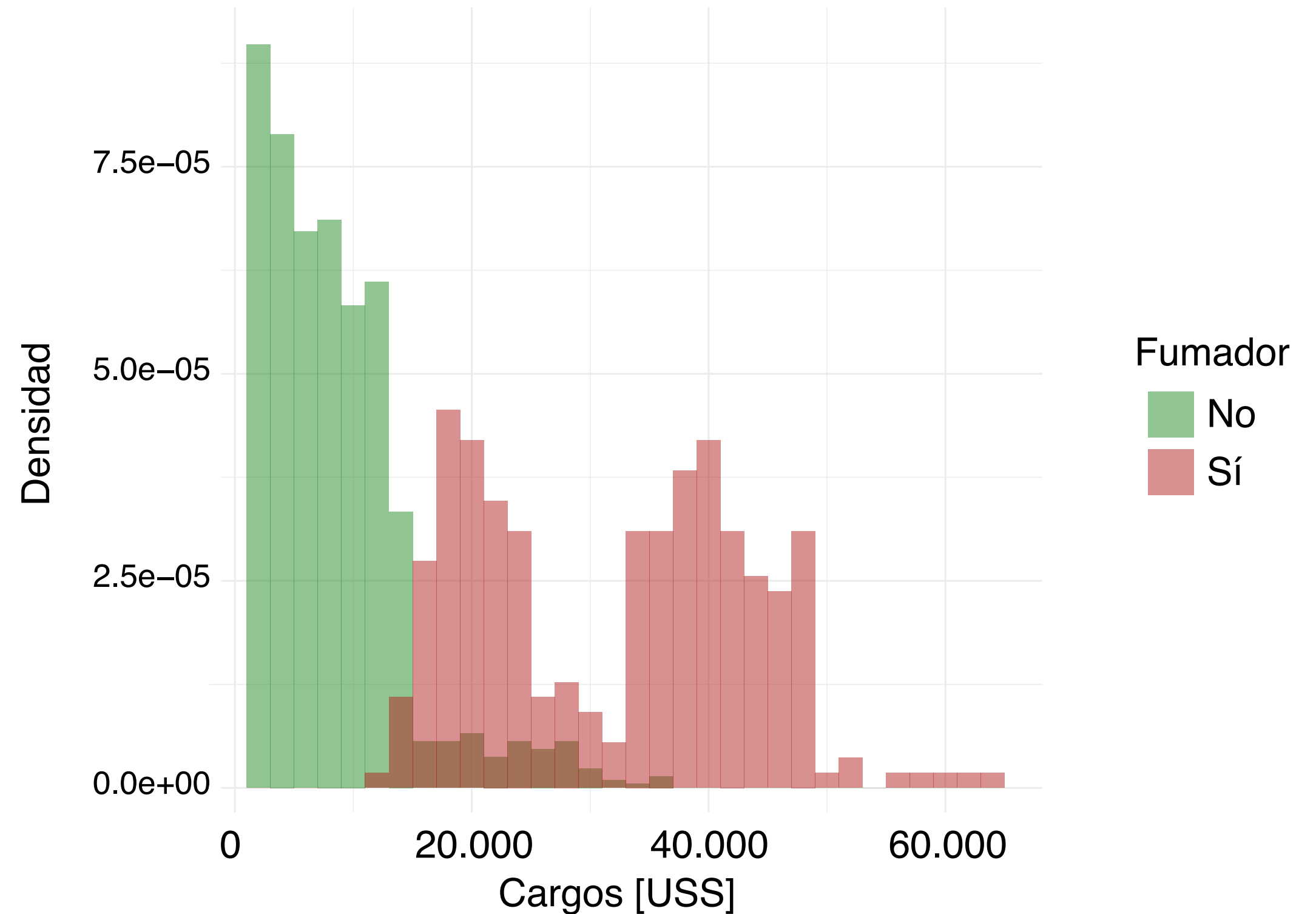
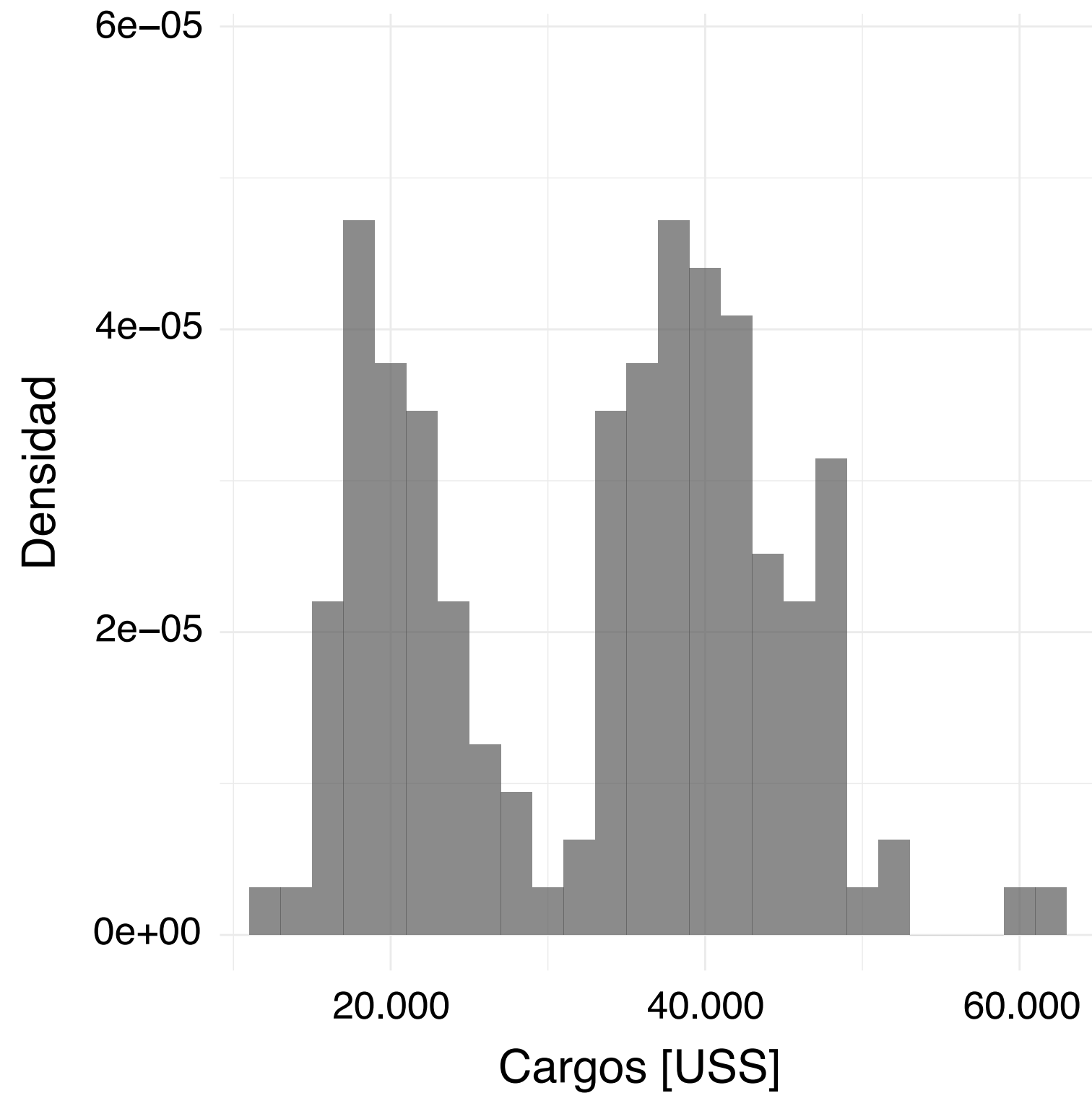


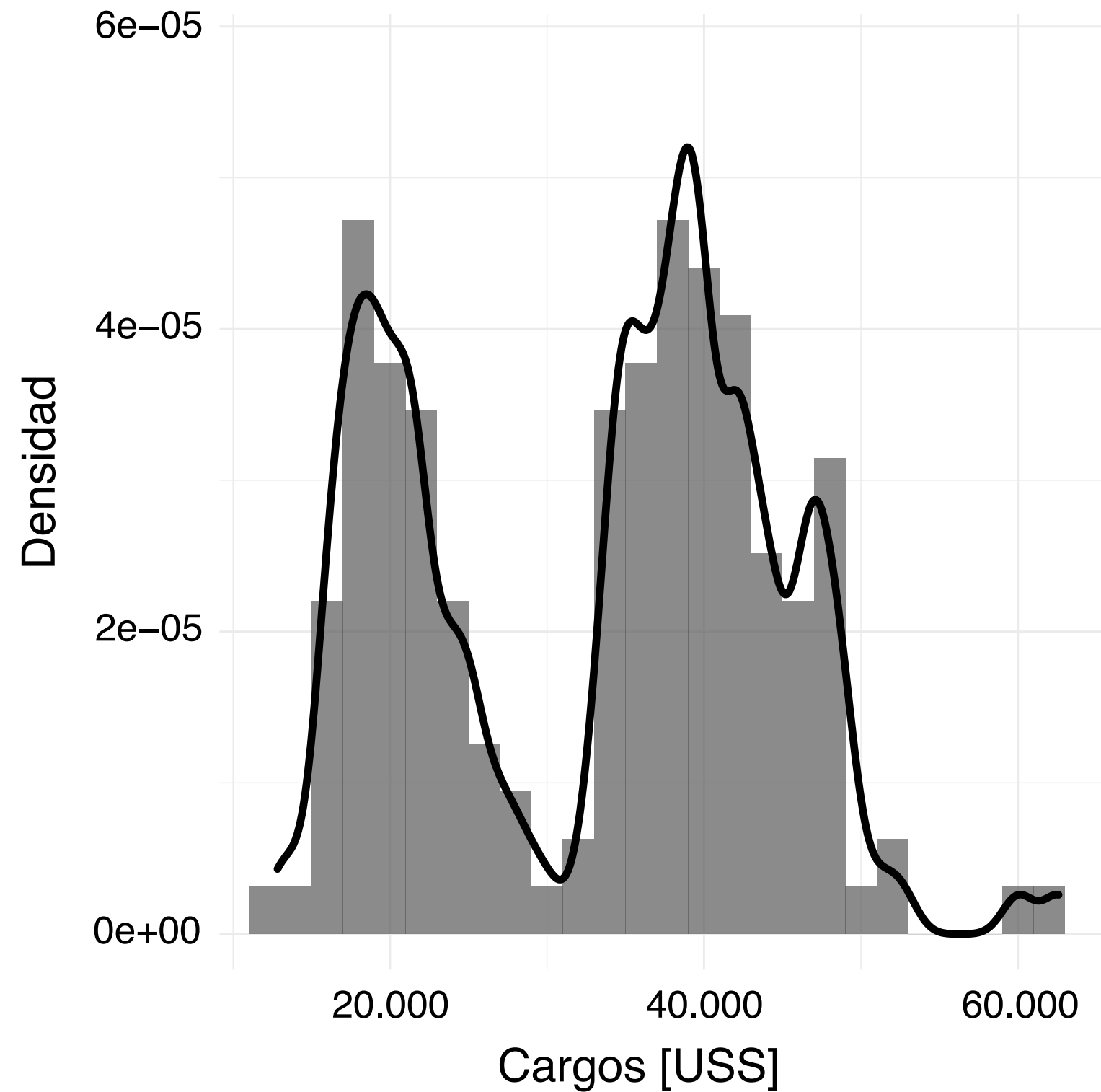
gráfico de densidad

```
geom_histogram(  
  y = ..density..)
```



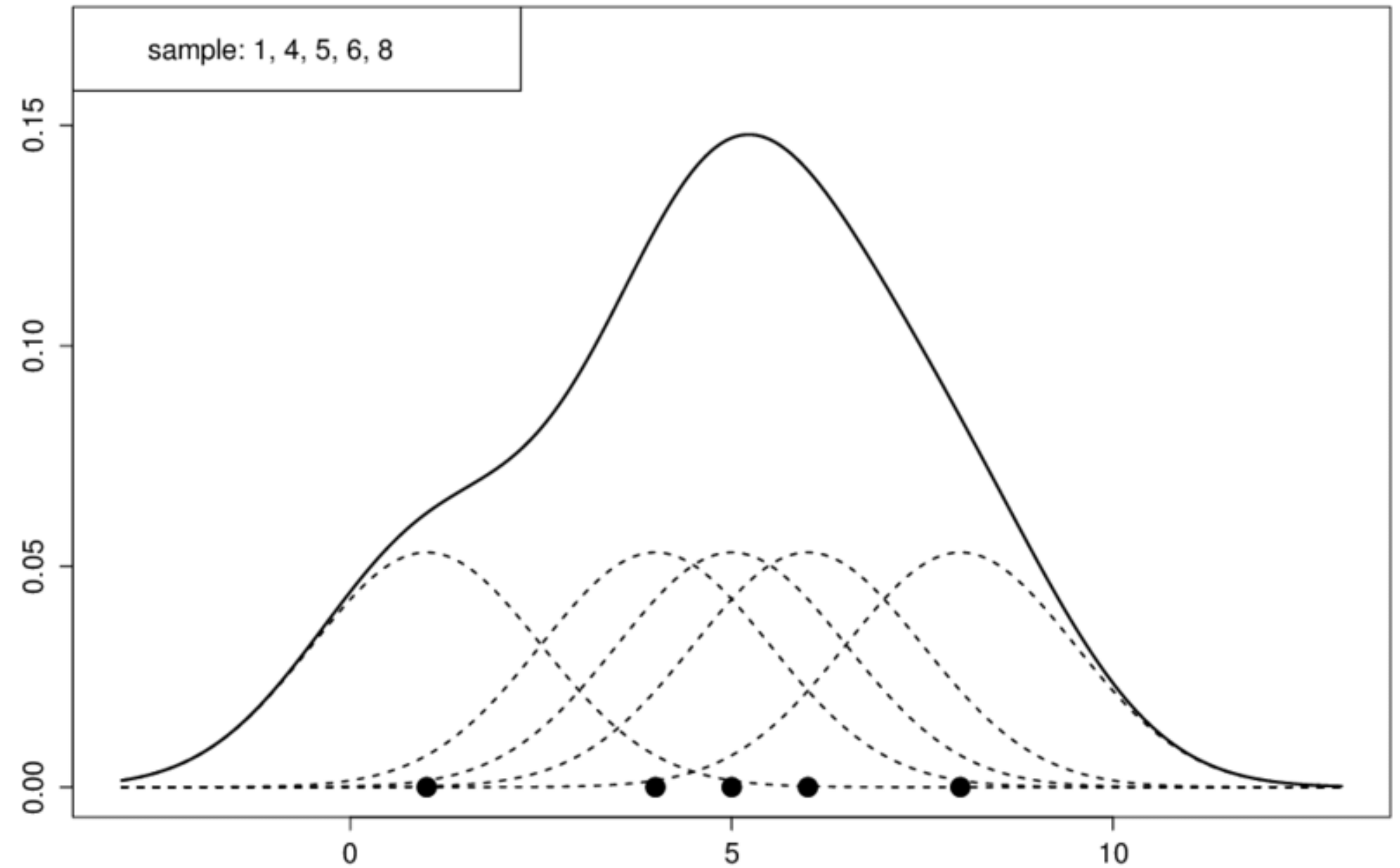
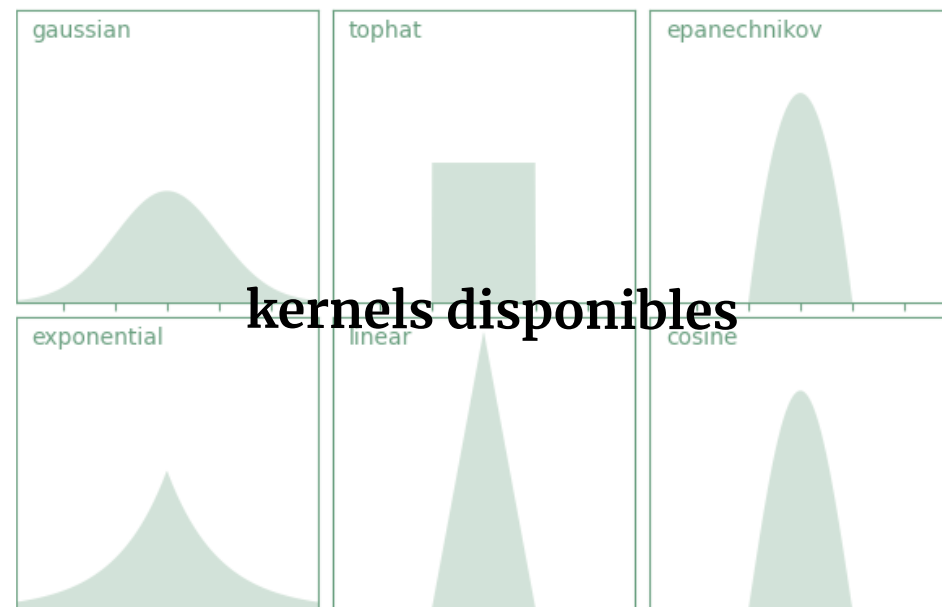
estimación de densidad con kernel

```
geom_histogram(  
  y = ..density..  
) +  
geom_density()
```



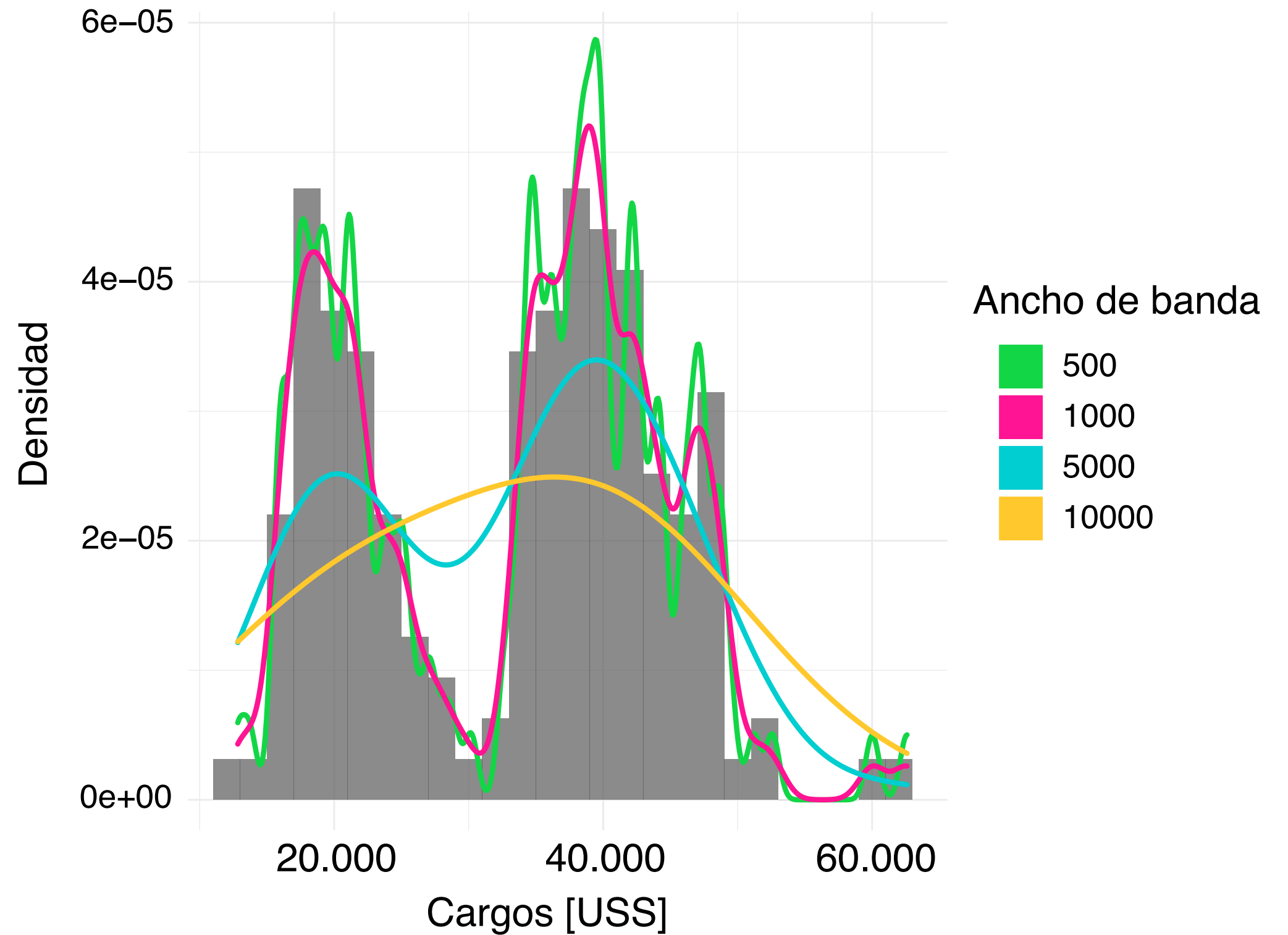
estimación de densidad con kernel

kernel gaussiano



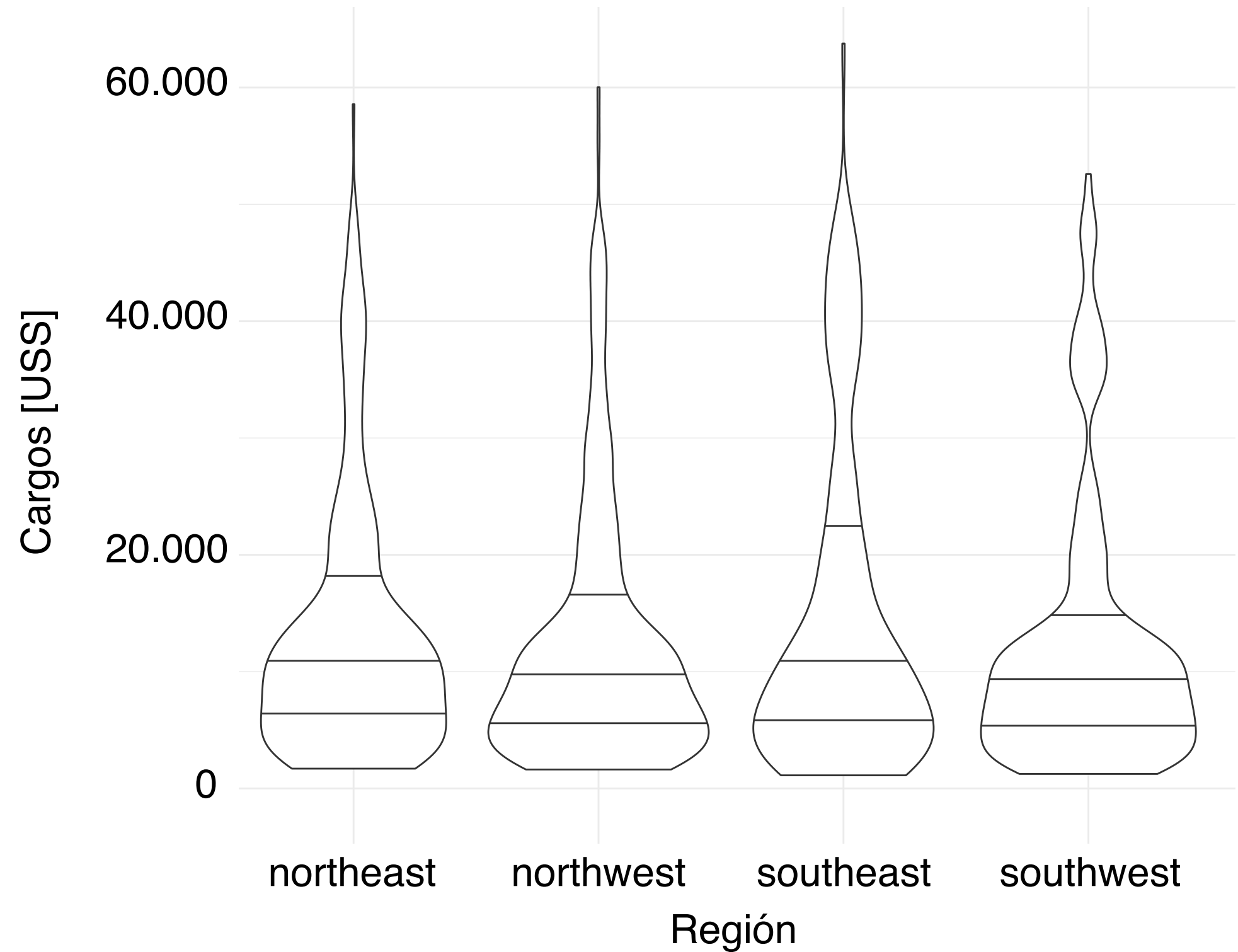
ancho de banda (bandwidth)

`geom_histogram(
 bw =)`



violinplot

```
geom_violin(  
  draw_quantiles=....)
```



Características de estas nuevas representaciones

- Visión más completa de la distribución
- Menos compacta que las métricas de resumen

