

## Continuous Variables (Chapter 3)

# Continuous Variables

We're looking for features such as:

- Asymmetry
- Outliers
- Multimodality
- Gaps
- Heaping / Rounding
- Impossibilities / Errors

primary tool: histograms

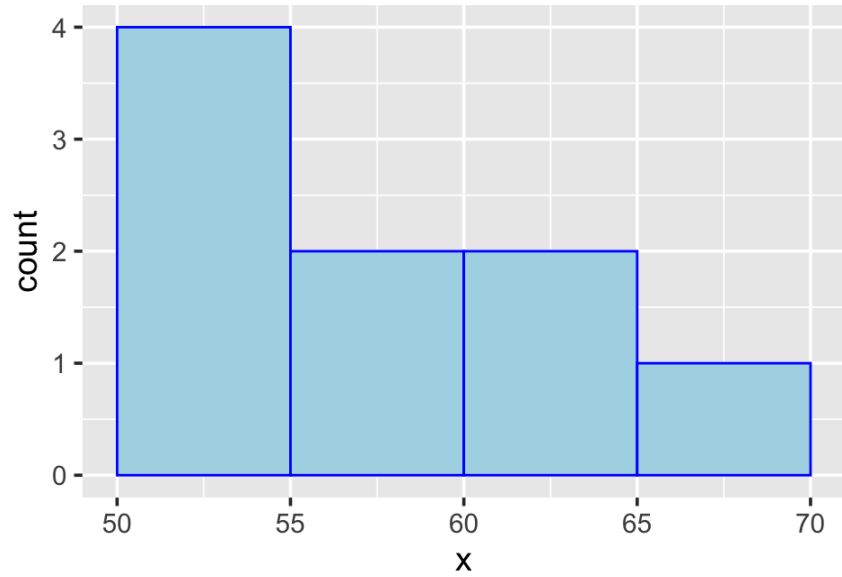
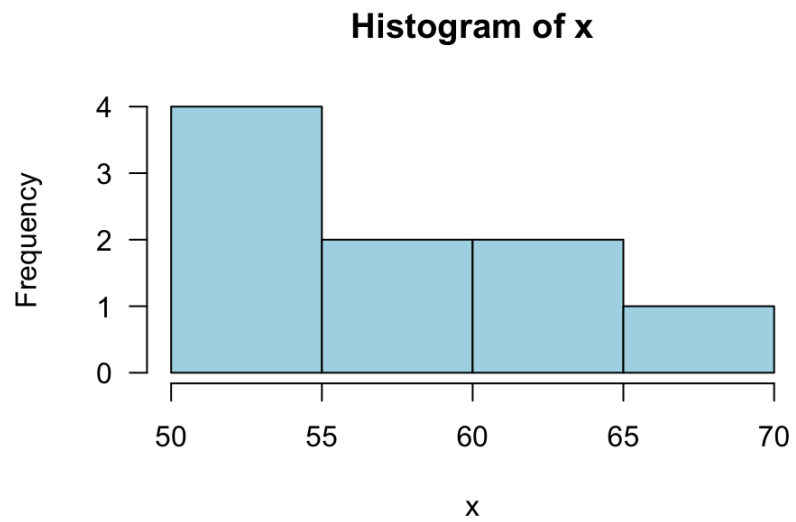
# How are histograms created?

Draw a histogram on paper of the following data.

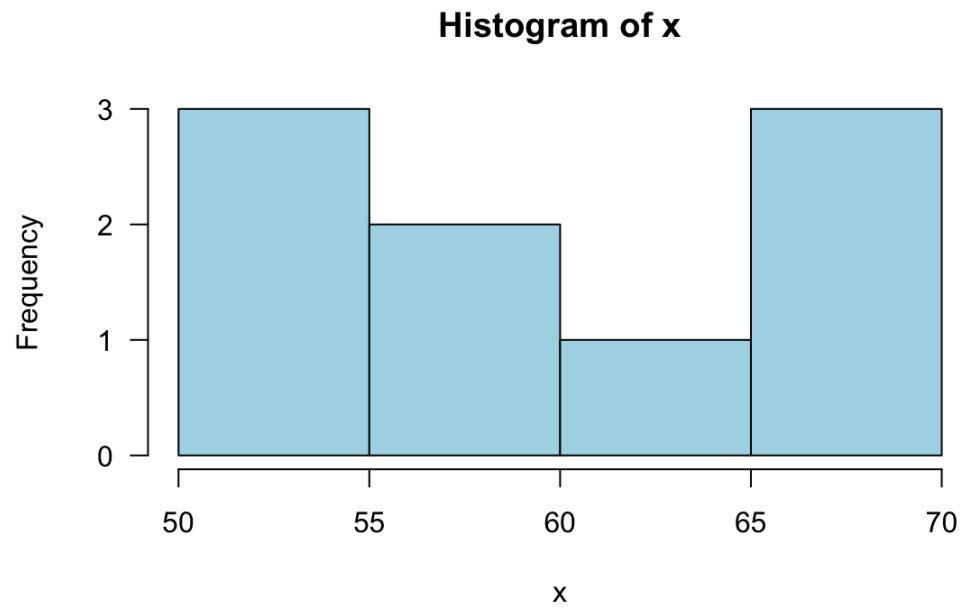
(use binwidth = 5)

50, 51, 53, 55, 56, 60, 65, 65, 68

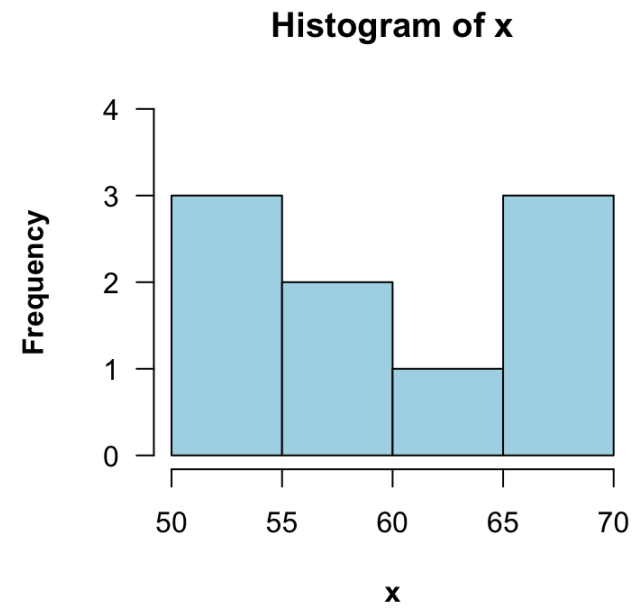
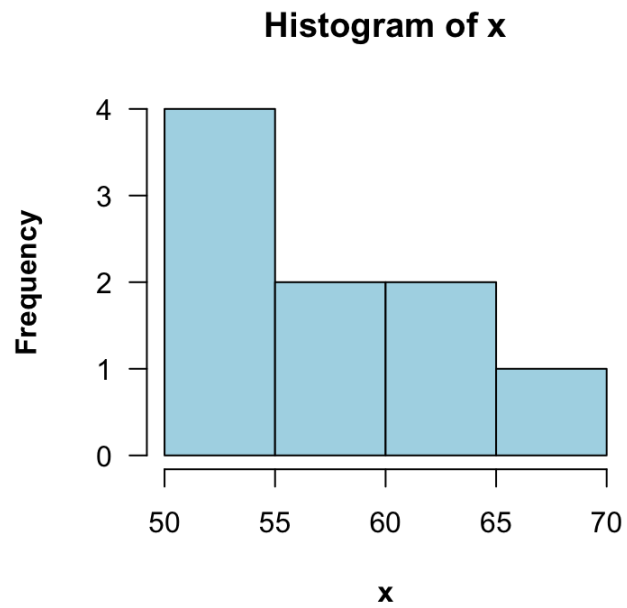
# How are histograms created?



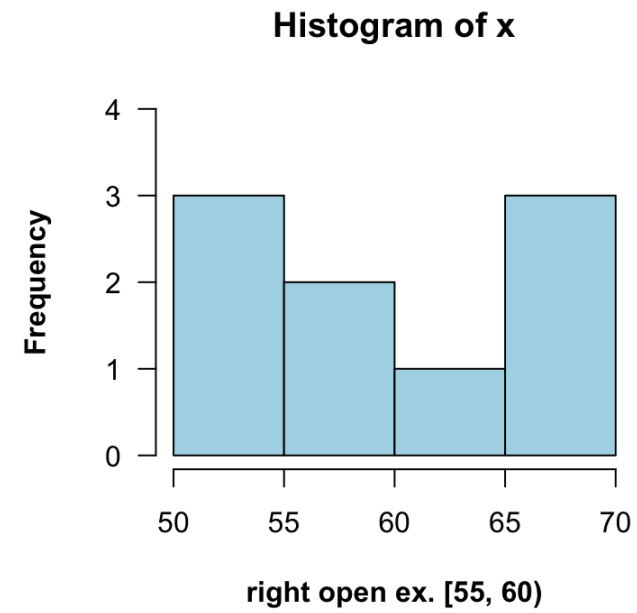
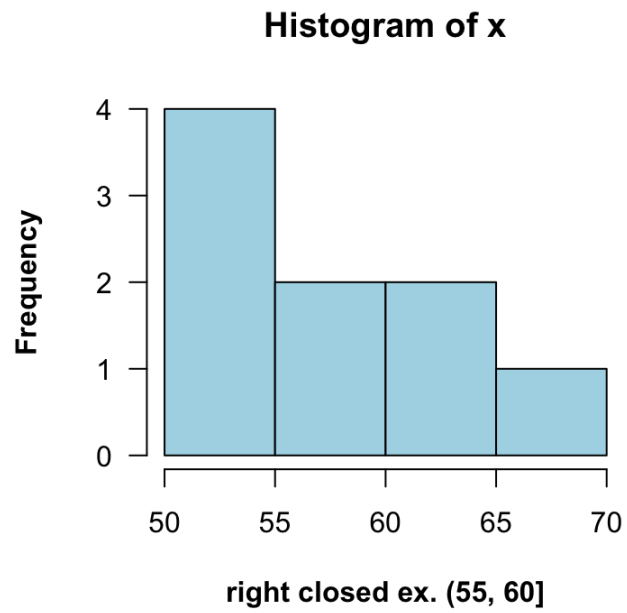
# How are histograms created?



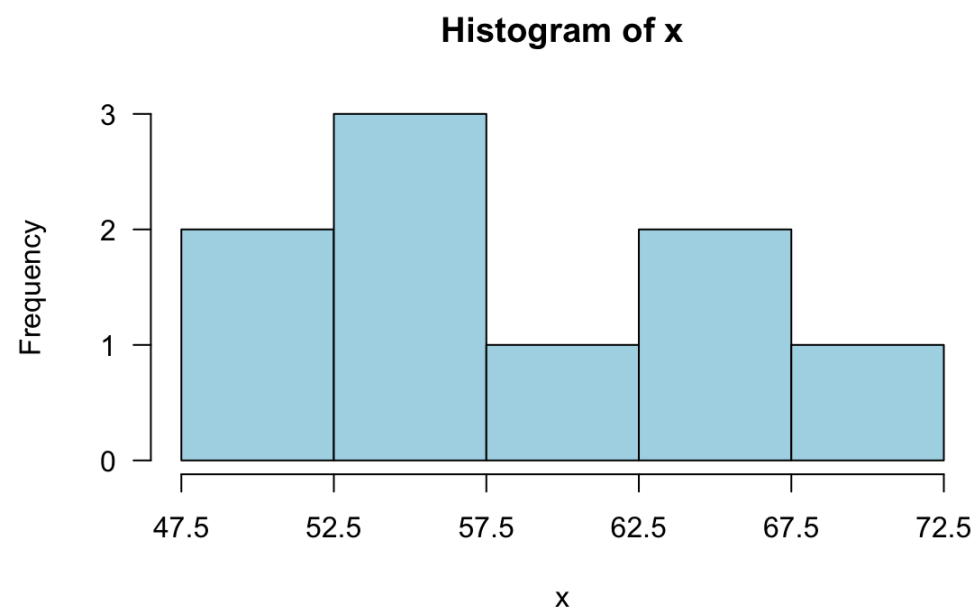
# What is causing the difference?



# Bin boundaries

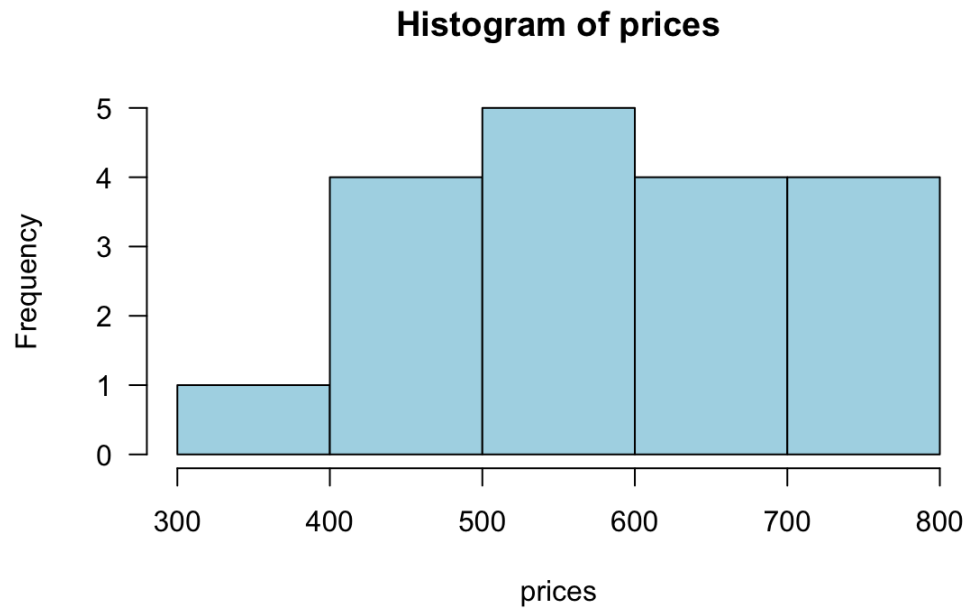


# Bin boundaries



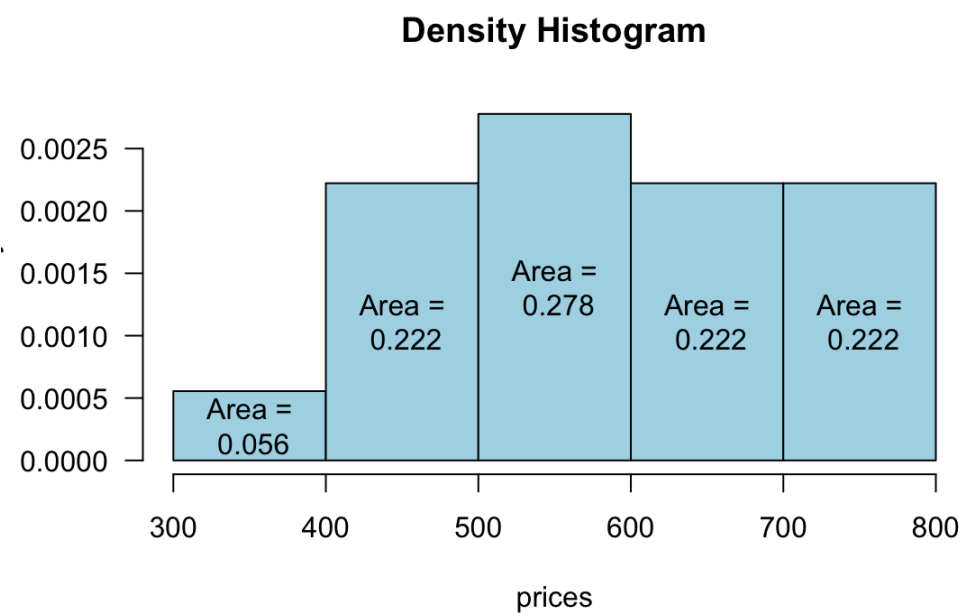


# Frequency (count) histogram



Prices of one-bedroom apartments in Morningside Heights (zip 10027) in \$1000k

# Density Histogram



# Count, Relative Frequency, Density

| Bin     | Count | Relative<br>Frequency | Binwidth | Density |
|---------|-------|-----------------------|----------|---------|
| 300-400 | 1     | .056                  | 100      | .00056  |
| 400-500 | 4     | .222                  | 100      | .00222  |
| 500-600 | 5     | .278                  | 100      | .00278  |
| 600-700 | 4     | .222                  | 100      | .00222  |
| 700-800 | 4     | .222                  | 100      | .00222  |

$\Sigma$  (bar areas) = 1

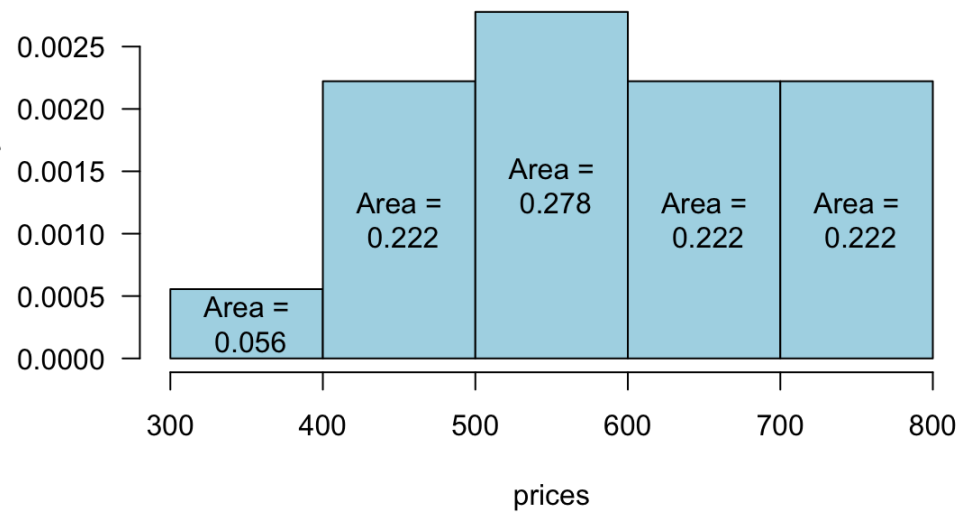
area of one bar =

relative frequency of that bar =

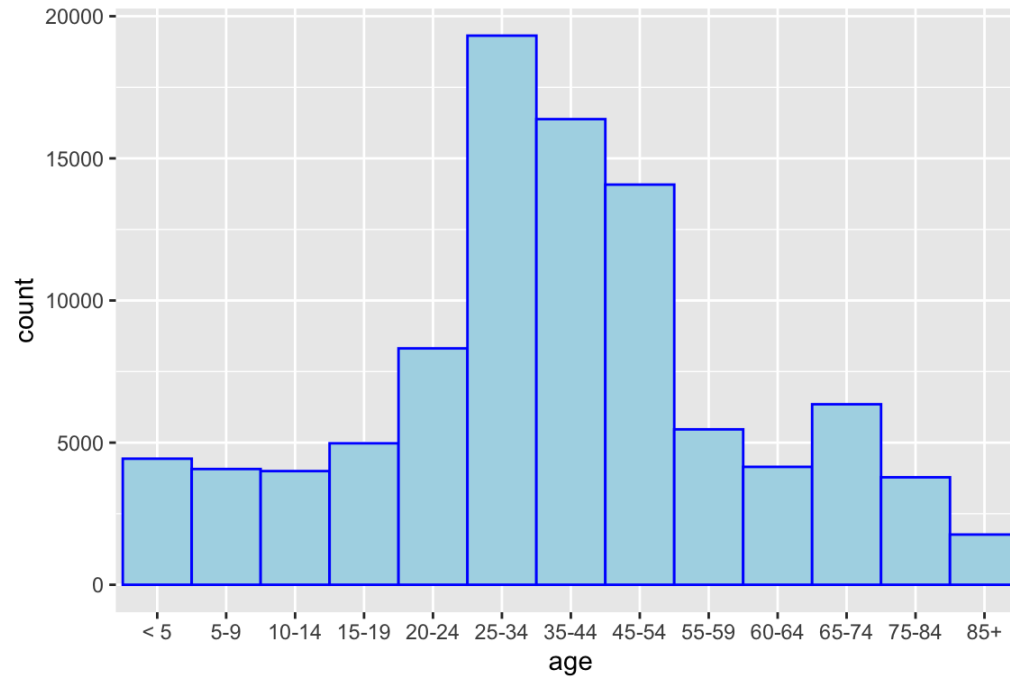
density \* binwidth

density = relative frequency /  
binwidth

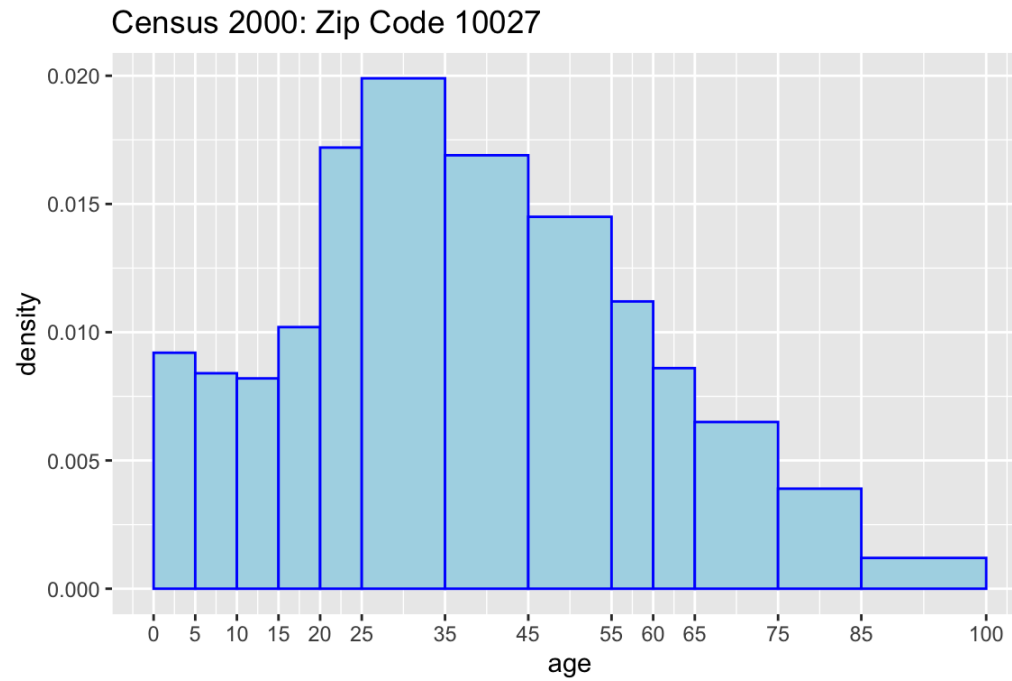
**Density Histogram**



# What's wrong with this histogram?



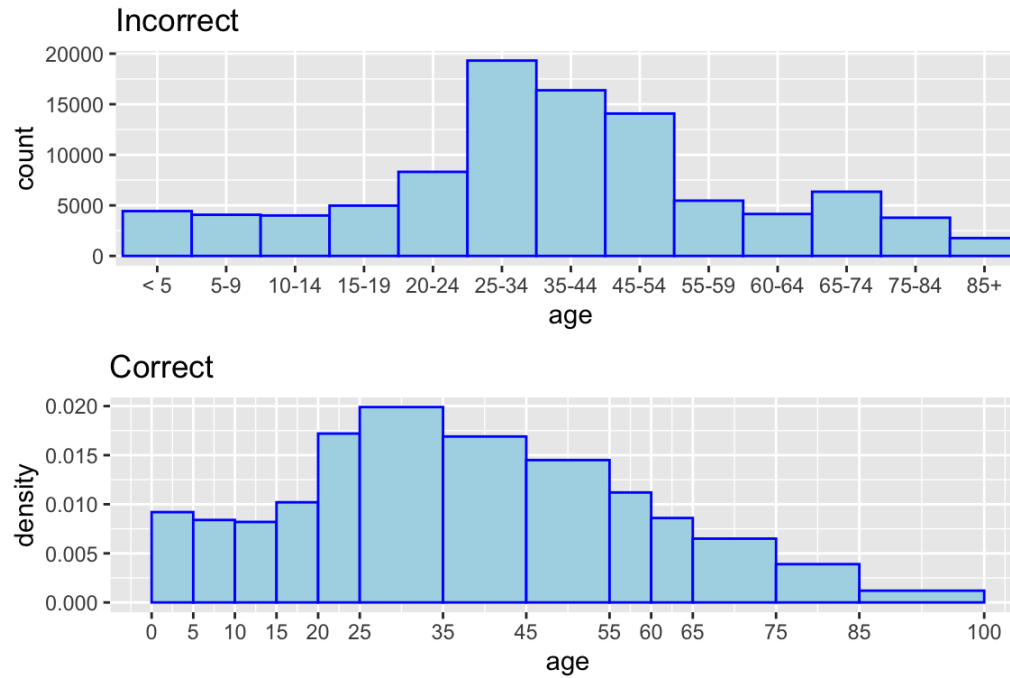
# Density histogram with unequal bin (or class) widths



$$\text{Density} = \text{RelFreq} / \text{Binwidth}$$

| <b>Class</b> | <b>Frequency</b> | <b>RelFreq</b> | <b>Binwidth</b> | <b>Density</b> |
|--------------|------------------|----------------|-----------------|----------------|
| < 5          | 4435             | 0.046          | 5               | 0.009          |
| 5-9          | 4072             | 0.042          | 5               | 0.008          |
| 10-14        | 3999             | 0.041          | 5               | 0.008          |
| 15-19        | 4977             | 0.051          | 5               | 0.010          |
| 20-24        | 8316             | 0.086          | 5               | 0.017          |
| 25-34        | 19317            | 0.199          | 10              | 0.020          |
| 35-44        | 16380            | 0.169          | 10              | 0.017          |
| 45-54        | 14077            | 0.145          | 10              | 0.014          |
| 55-59        | 5467             | 0.056          | 5               | 0.011          |
| 60-64        | 4148             | 0.043          | 5               | 0.009          |
| 65-74        | 6350             | 0.065          | 10              | 0.007          |
| 75-84        | 3781             | 0.039          | 10              | 0.004          |
| 85+          | 1767             | 0.018          | 15              | 0.001          |

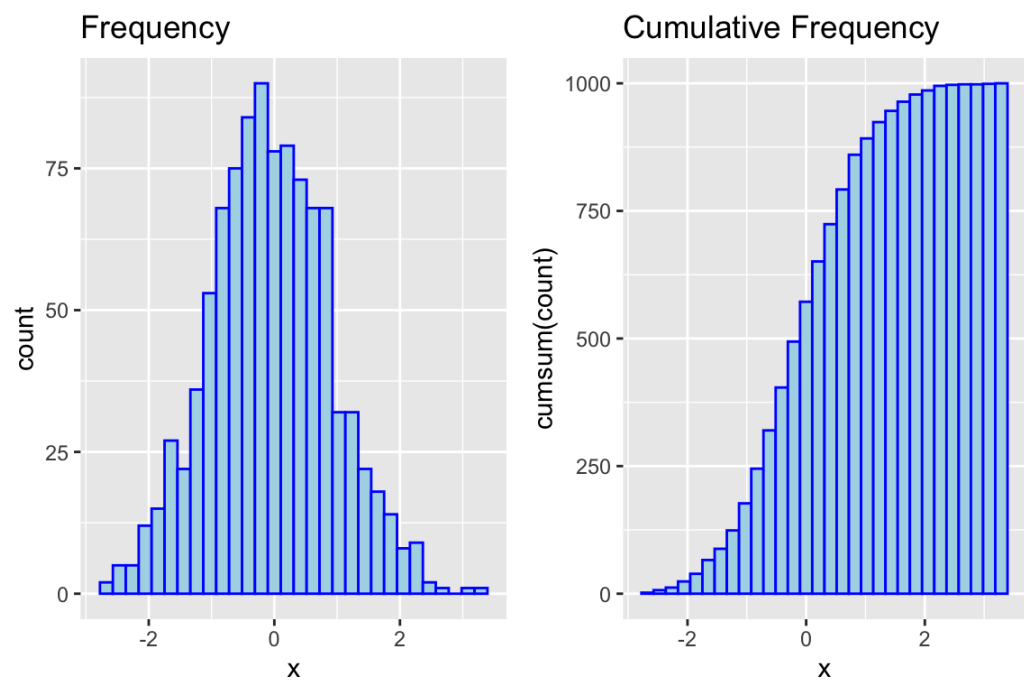
# Compare the histograms



Source: <https://factfinder.census.gov/>



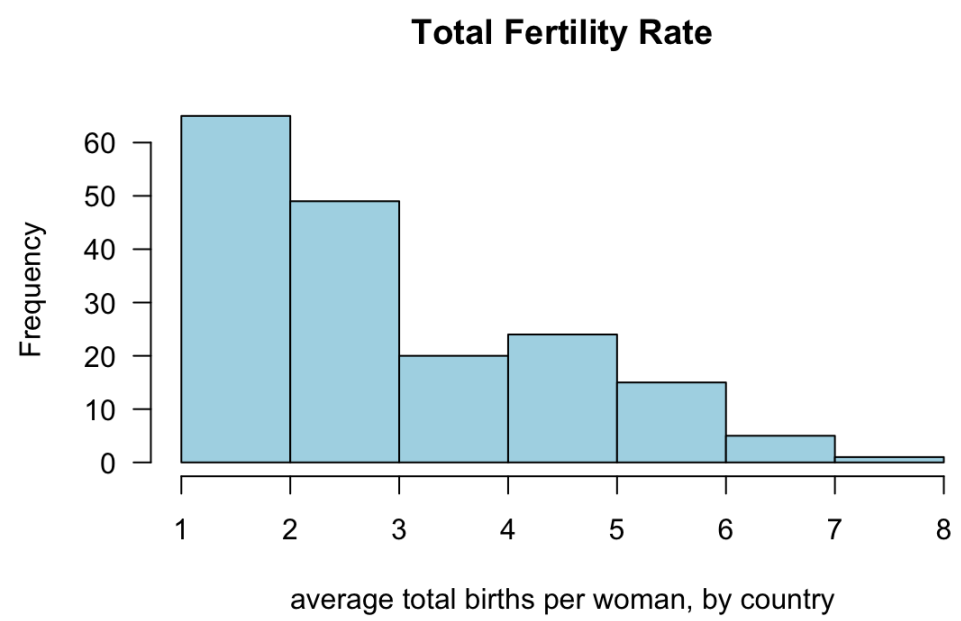
# Cumulative frequency histogram



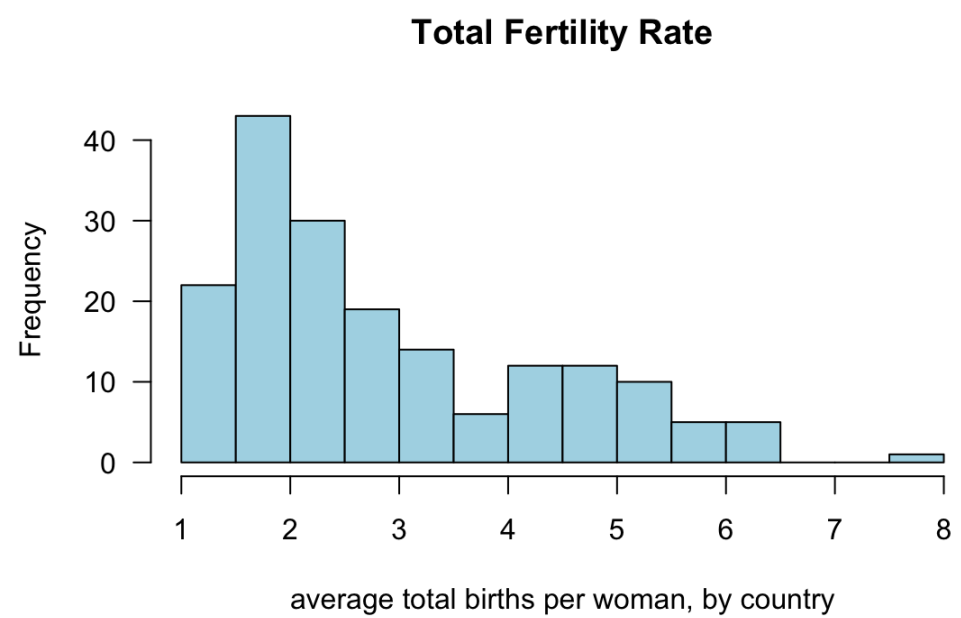
## Binwidth

'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

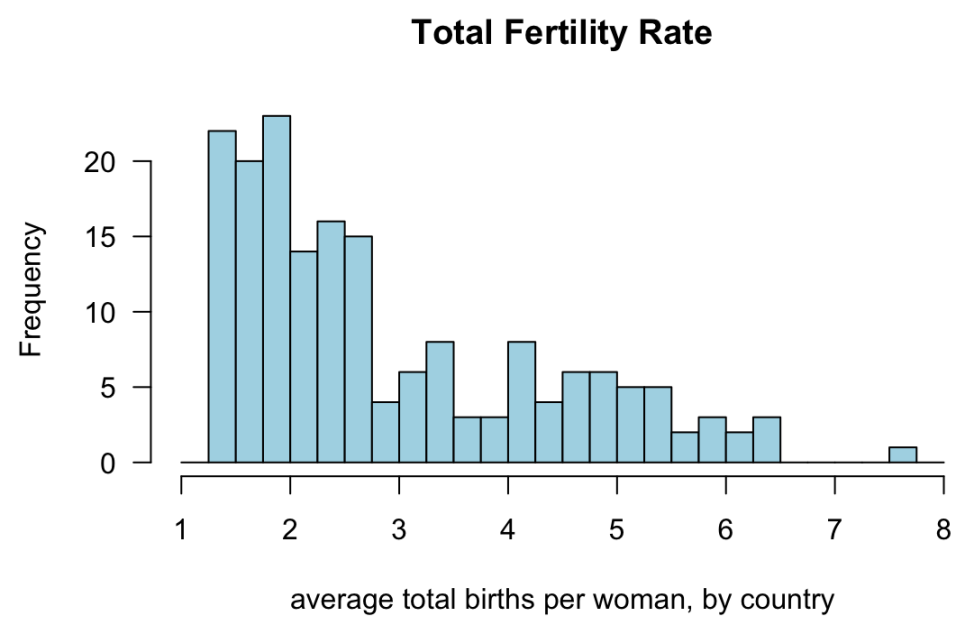
# Histograms



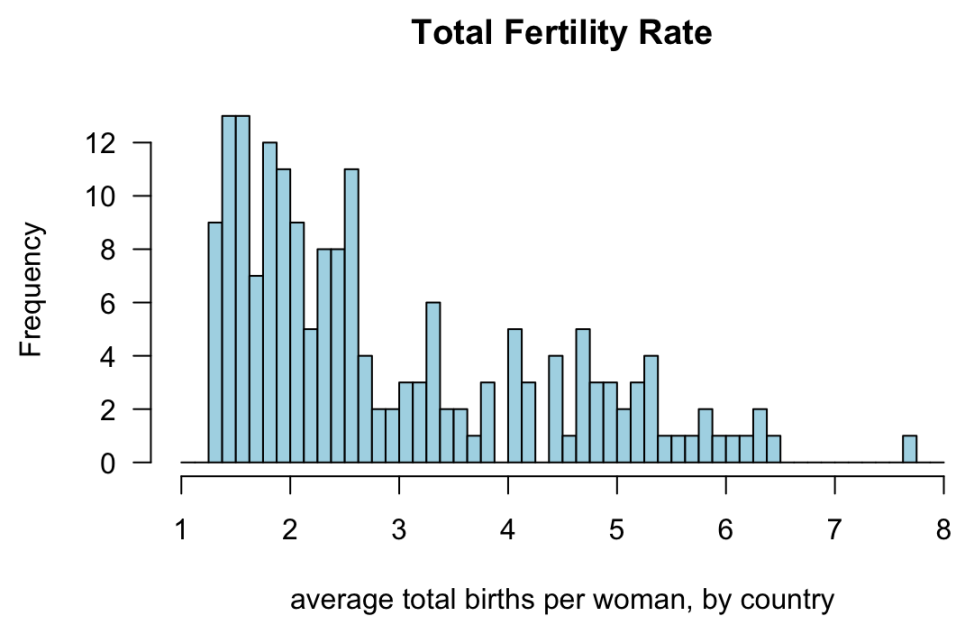
# Histograms



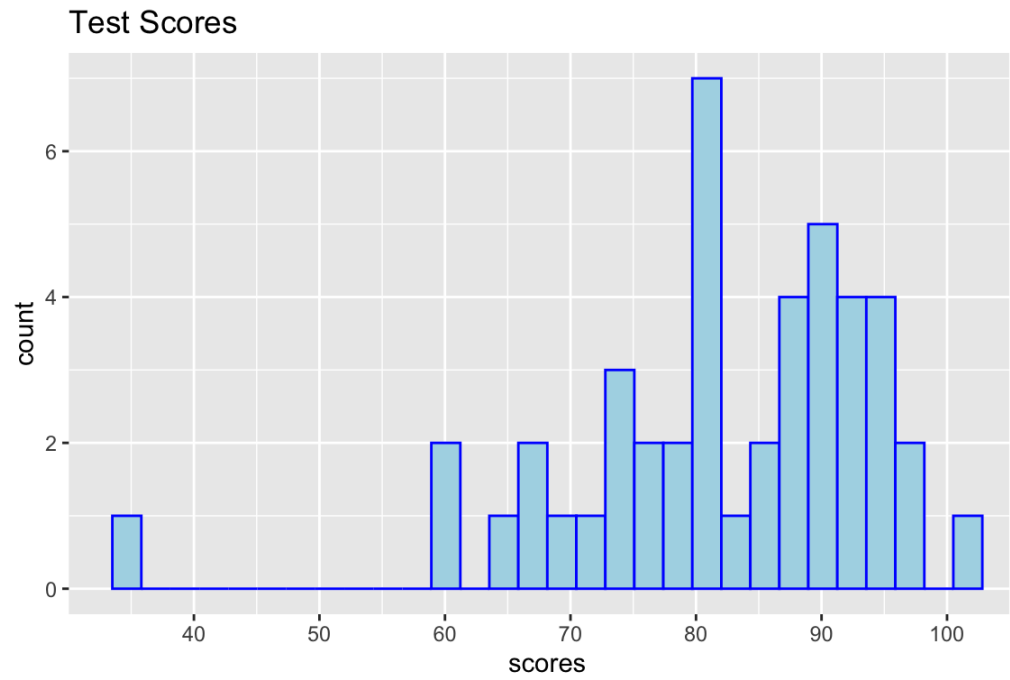
# Histograms



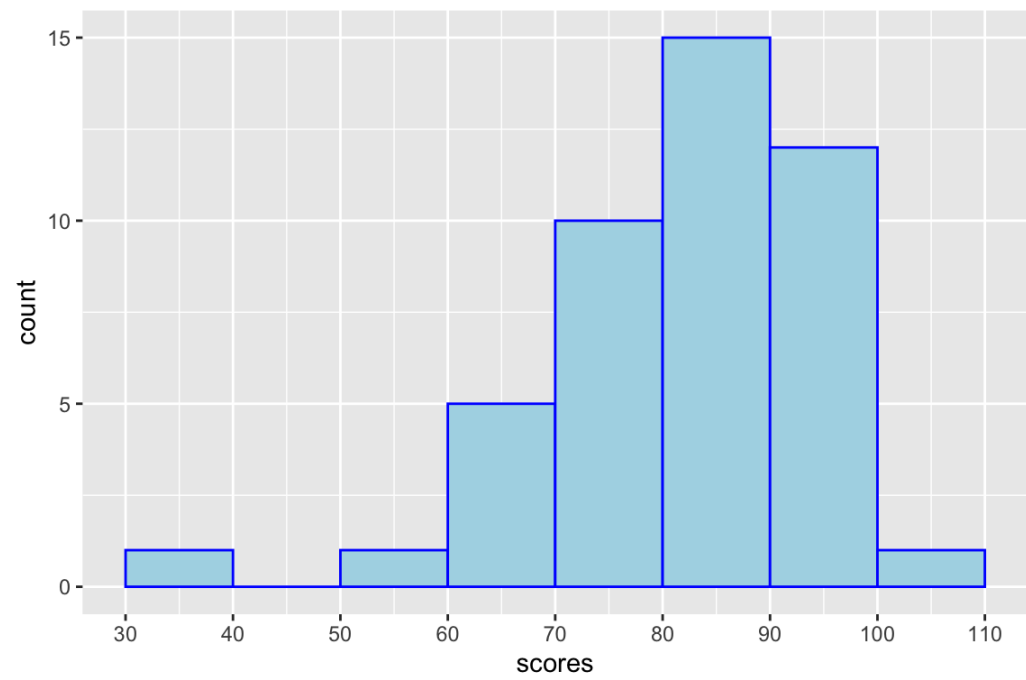
# Histograms



# Histograms



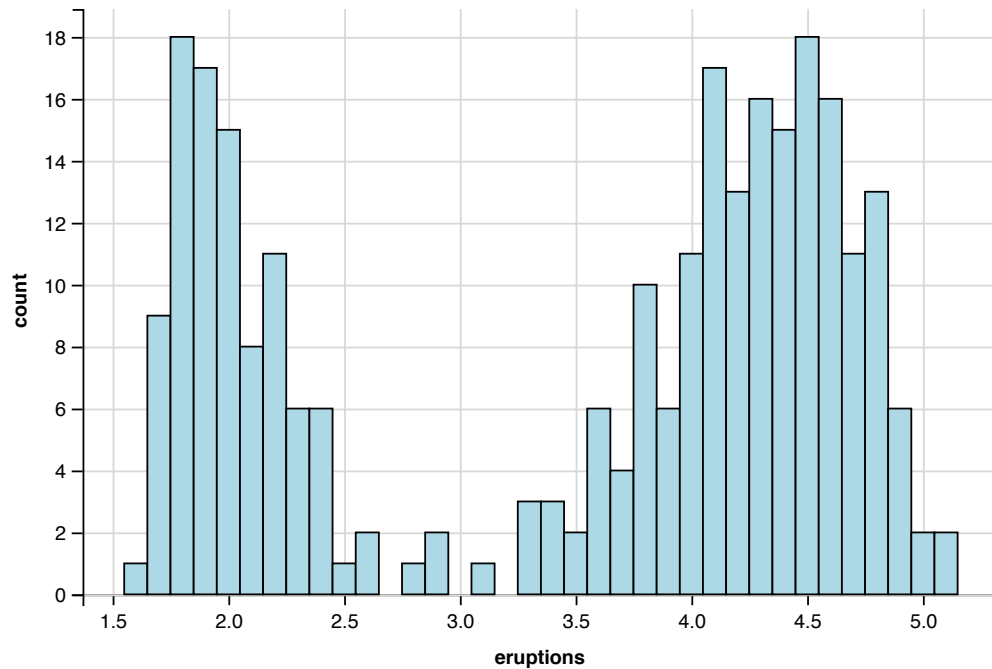
# Fewer bins





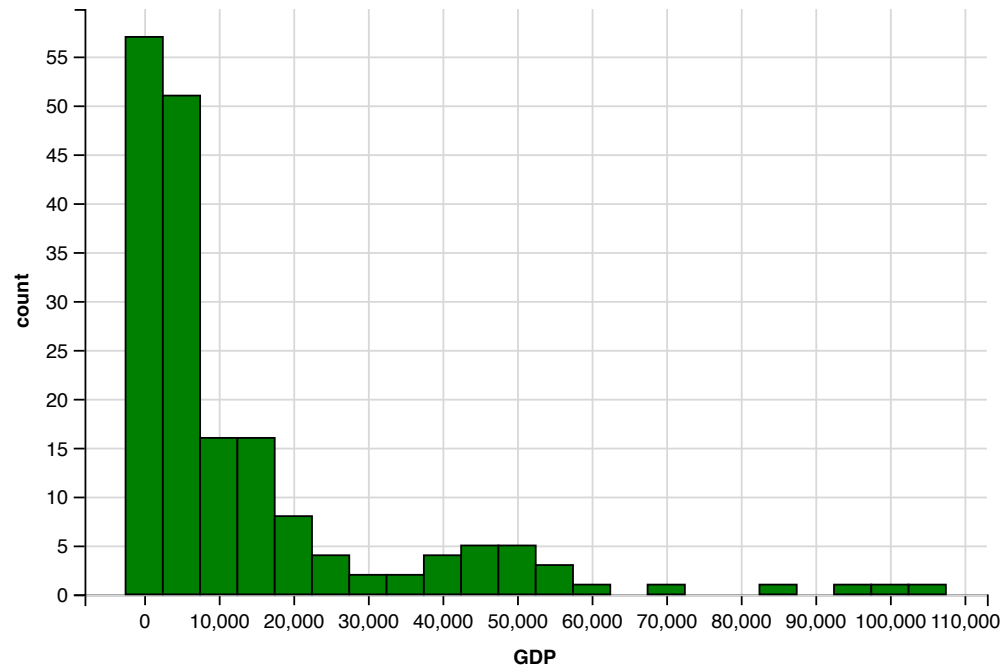
# Change binwidth interactively

```
## Warning: Can't output dynamic/interactive ggvis plots in a knitr document.  
## Generating a static (non-dynamic, non-interactive) version of the plot.
```



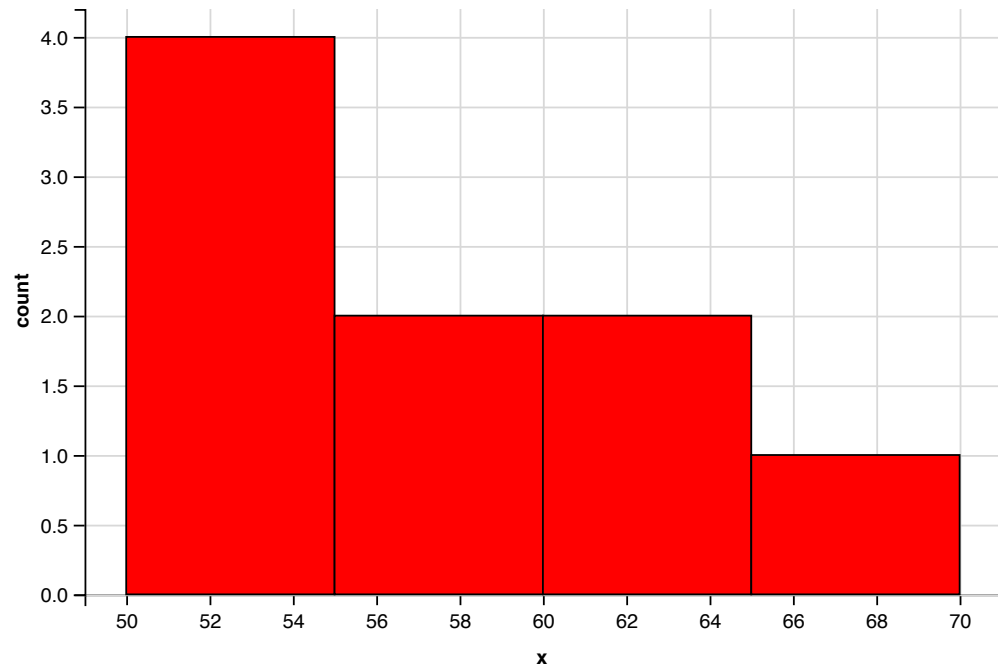
# GDP

```
## Warning: Can't output dynamic/interactive ggvis plots in a knitr document.
## Generating a static (non-dynamic, non-interactive) version of the plot.
```



# Center

```
## Warning: Can't output dynamic/interactive ggvis plots in a knitr document.  
## Generating a static (non-dynamic, non-interactive) version of the plot.
```



# Center (with data values shown)

```
## Warning: Can't output dynamic/interactive ggvis plots in a knitr document.  
## Generating a static (non-dynamic, non-interactive) version of the plot.
```

