Data visualization and reshaping

Lecture 3

IFT6758, Fall 2020

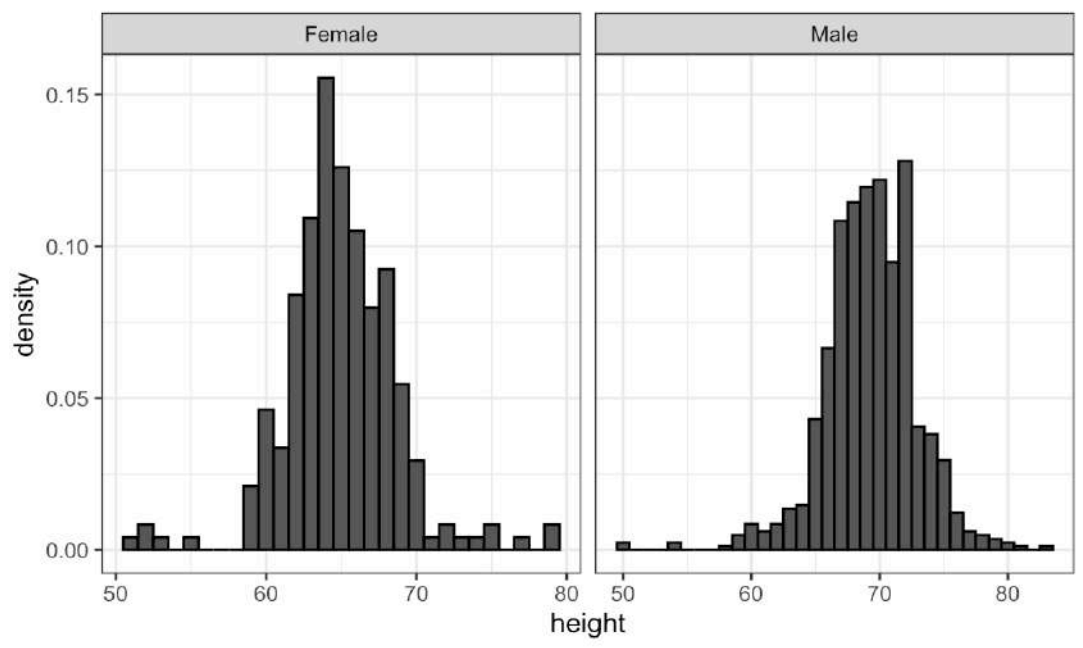
Principles of Data Visualization: **4. Ease comparisons**

Principles of Data Visualization: **4. Ease comparisons**

- **(i) Use common axes**

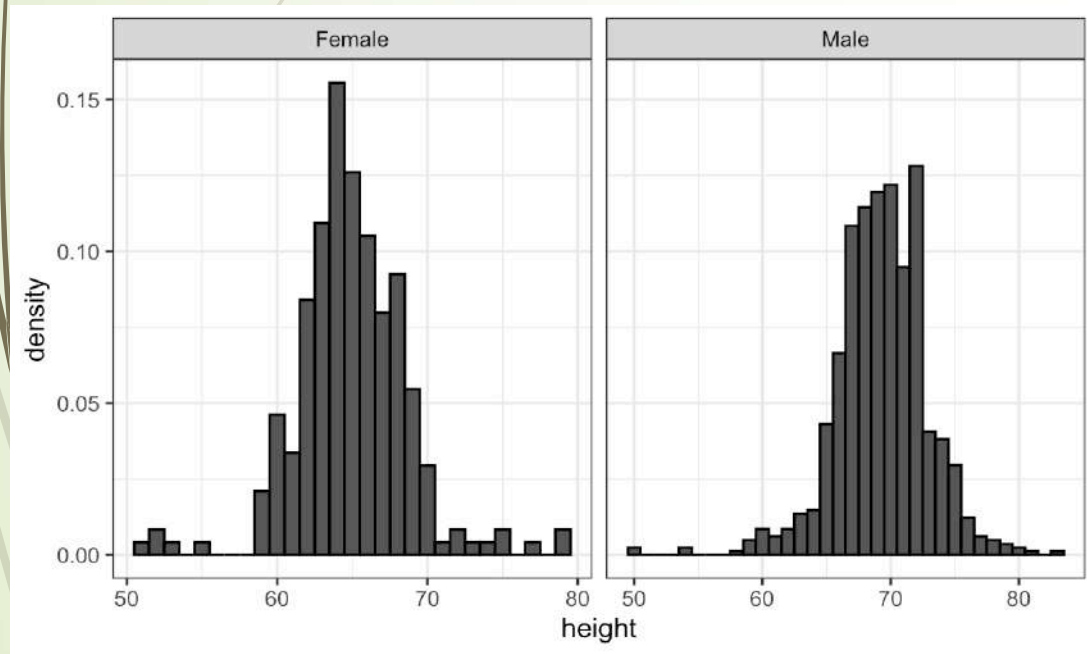
Principles of Data Visualization: **4. Ease comparisons**

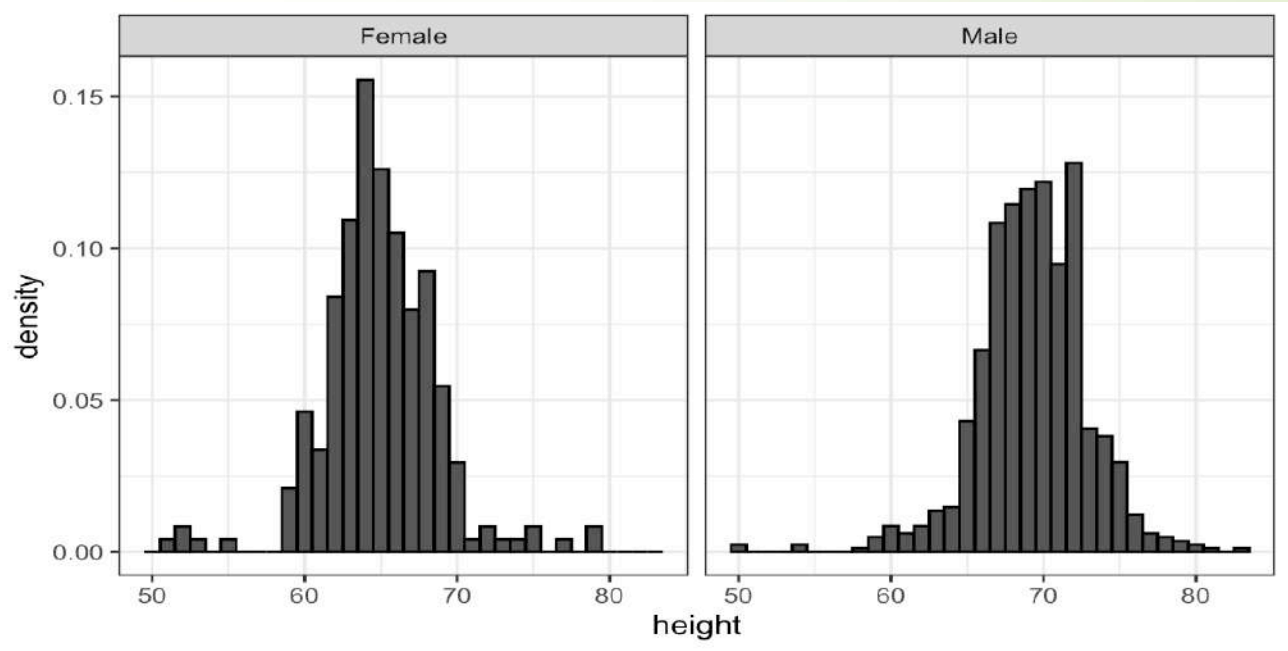
- **(i) Use common axes**



Principles of Data Visualization: **4. Ease comparisons**

- **(i) Use common axes**





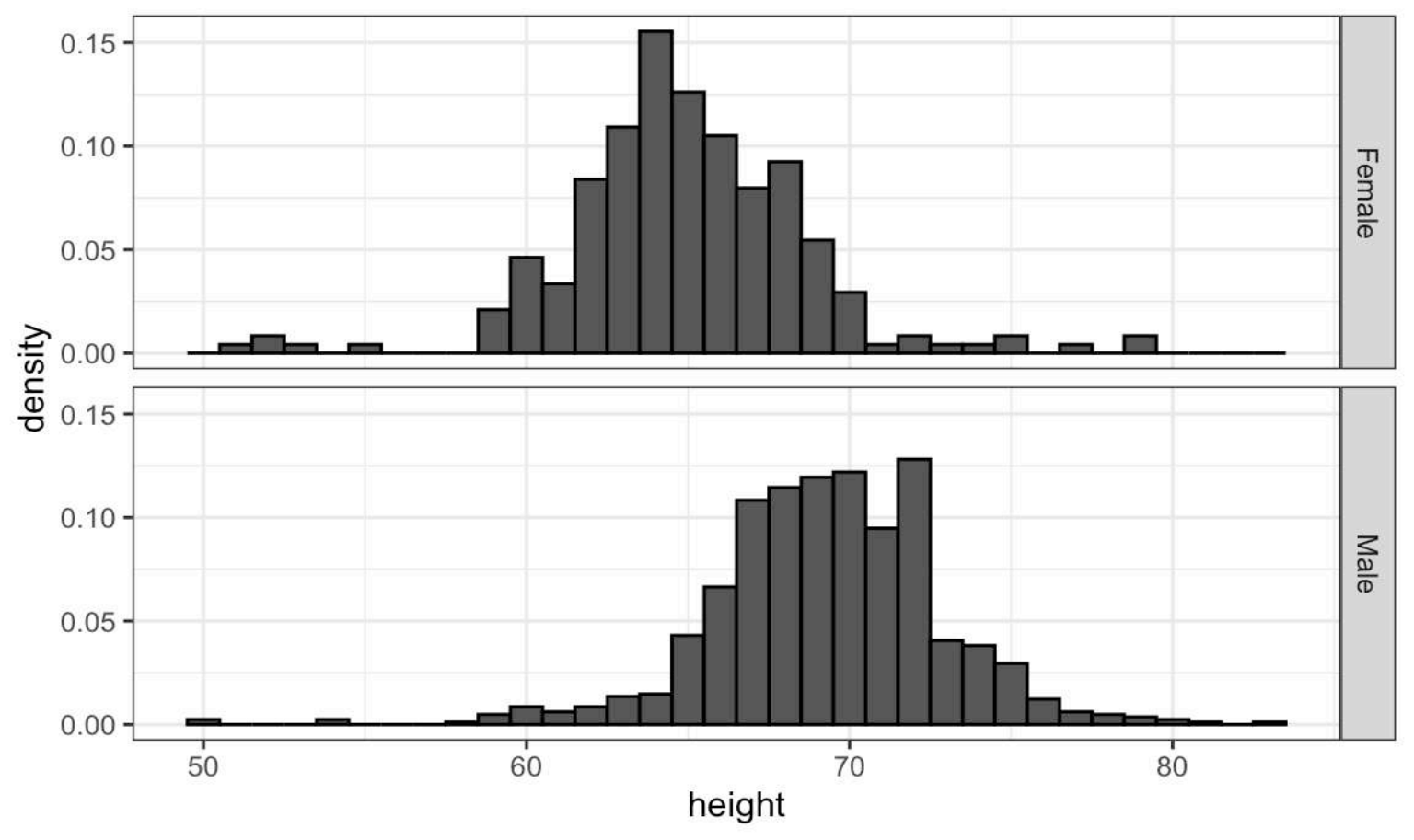
Principles of Data Visualization: **4. Ease comparisons**

Principles of Data Visualization: **4. Ease comparisons**

- **(ii) Align plots vertically to see horizontal changes and horizontally to see vertical changes**

Principles of Data Visualization: **4. Ease comparisons**

- **(ii) Align plots vertically to see horizontal changes and horizontally to see vertical changes**

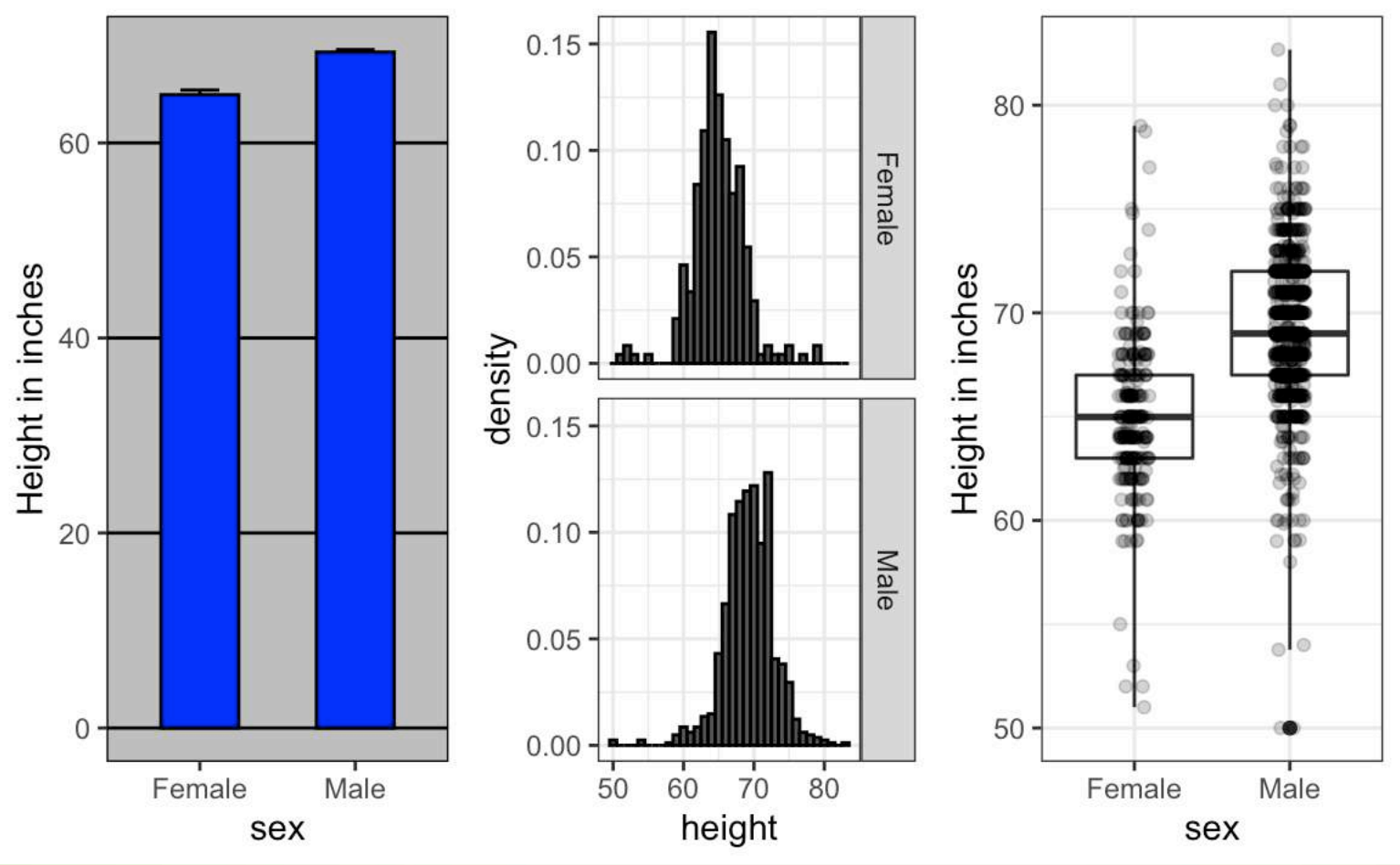


Principles of Data Visualization: **4. Ease comparisons**

- **(iii) Use appropriate representation to facilitate distribution and summary**

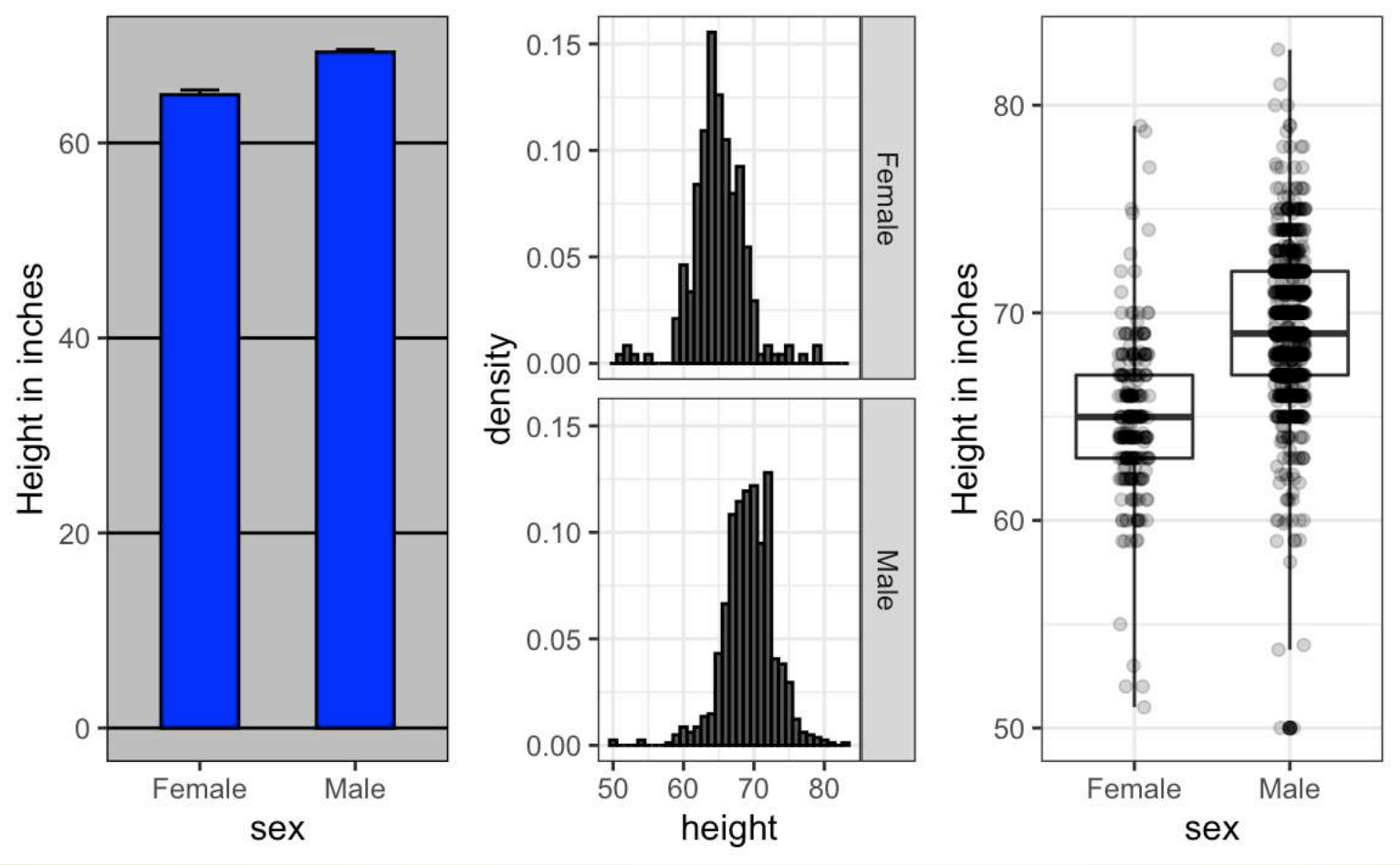
Principles of Data Visualization: **4. Ease comparisons**

- **(iii) Use appropriate representation to facilitate distribution and summary**



Principles of Data Visualization: **4. Ease comparisons**

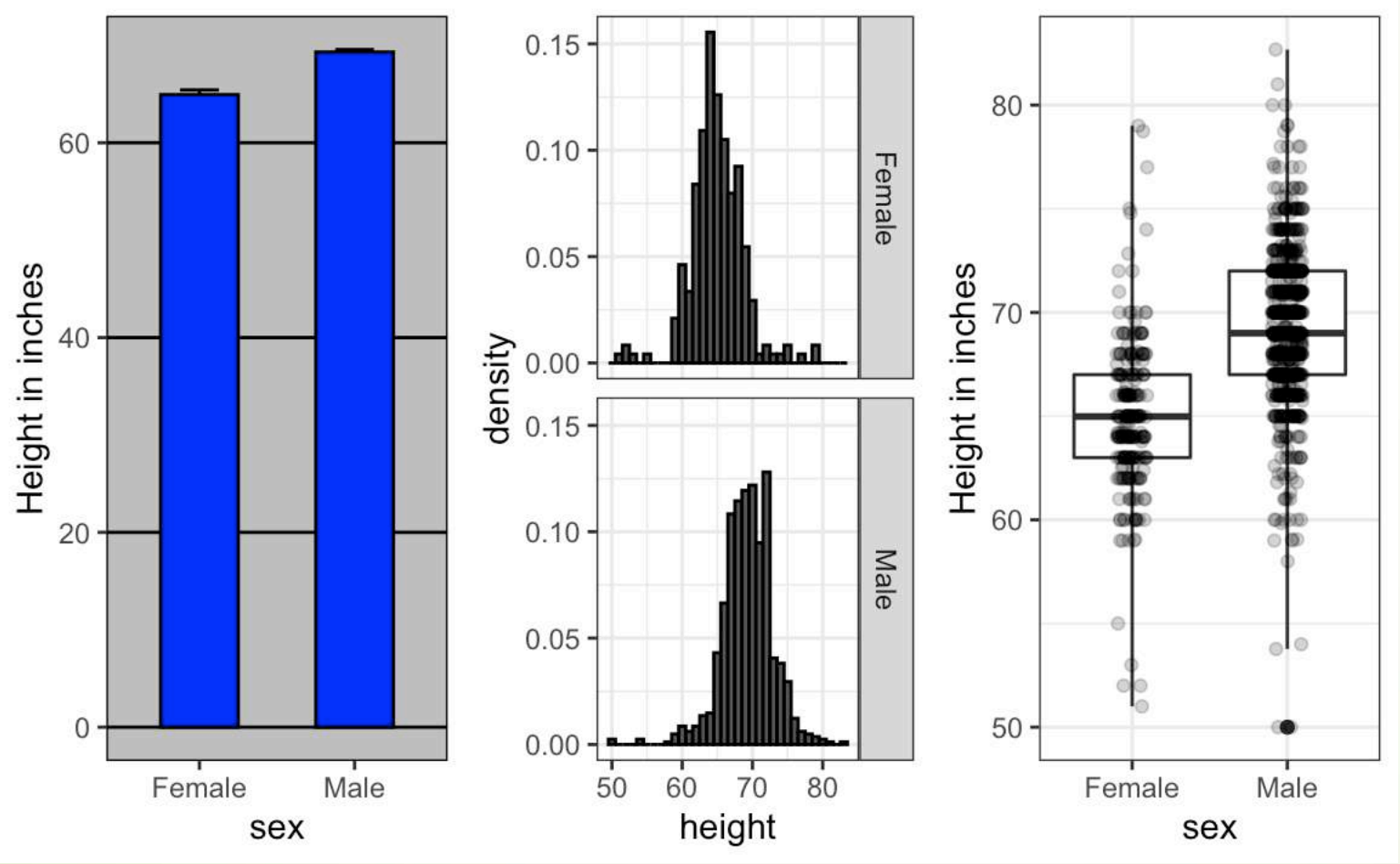
- **(iii) Use appropriate representation to facilitate distribution and summary**



matplotlib: pybeeswarm

Principles of Data Visualization: **4. Ease comparisons**

- **(iii) Use appropriate representation to facilitate distribution and summary**

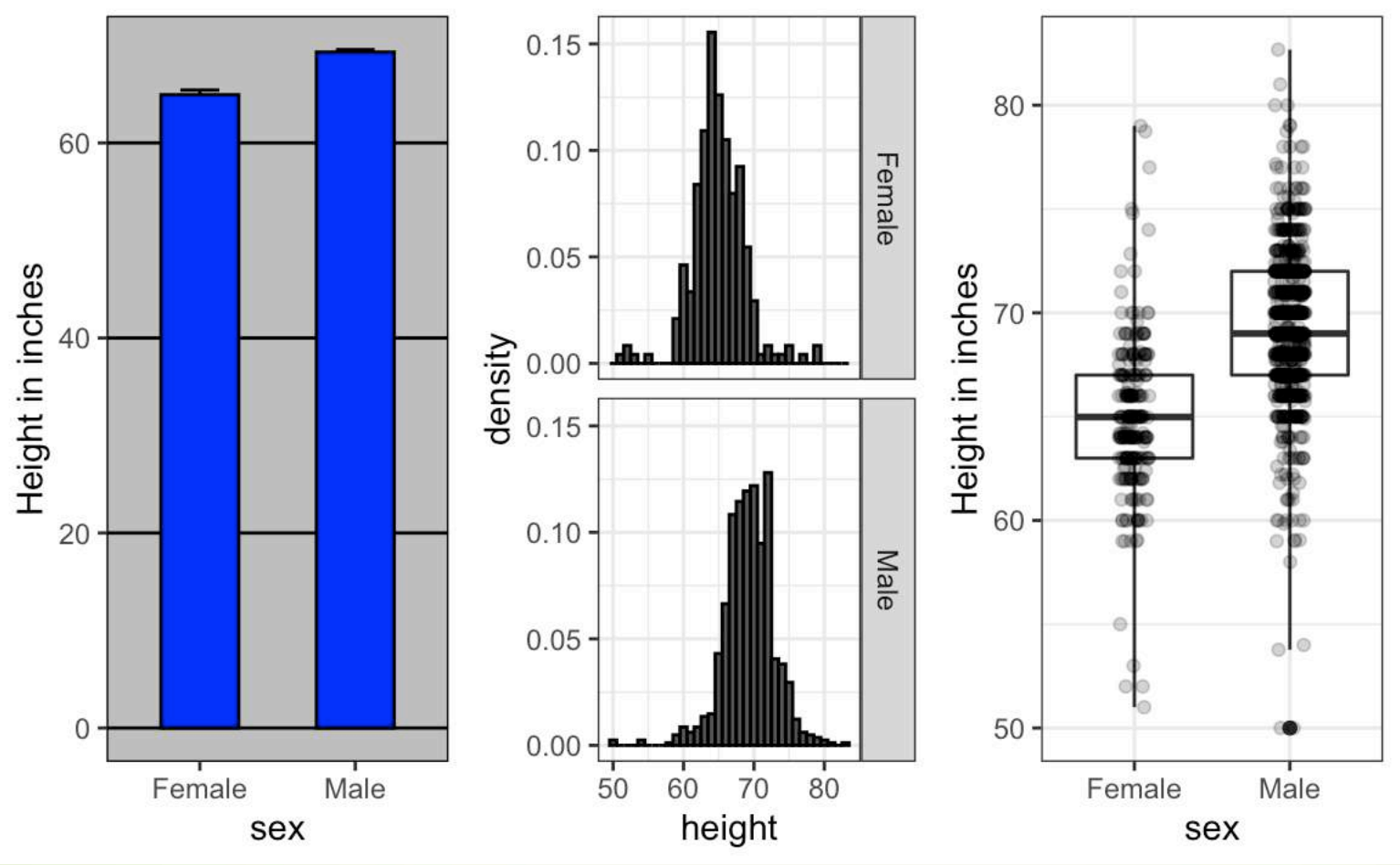


matplotlib: pybeeswarm

```
from beeswarm import *
import matplotlib.pyplot as plt
import numpy as np
d1 = np.random.uniform(low=-3, high=3, size=100)
d2 = np.random.normal(size=100)
```

Principles of Data Visualization: **4. Ease comparisons**

- **(iii) Use appropriate representation to facilitate distribution and summary**



matplotlib: pybeeswarm

```
from beeswarm import *
import matplotlib.pyplot as plt
import numpy as np
d1 = np.random.uniform(low=-3, high=3, size=100)
d2 = np.random.normal(size=100)
```

```
bs, ax = beeswarm([d1,d2], method="swarm", labels=["sample 1", "sample 2"], col=["blue","red"])
```

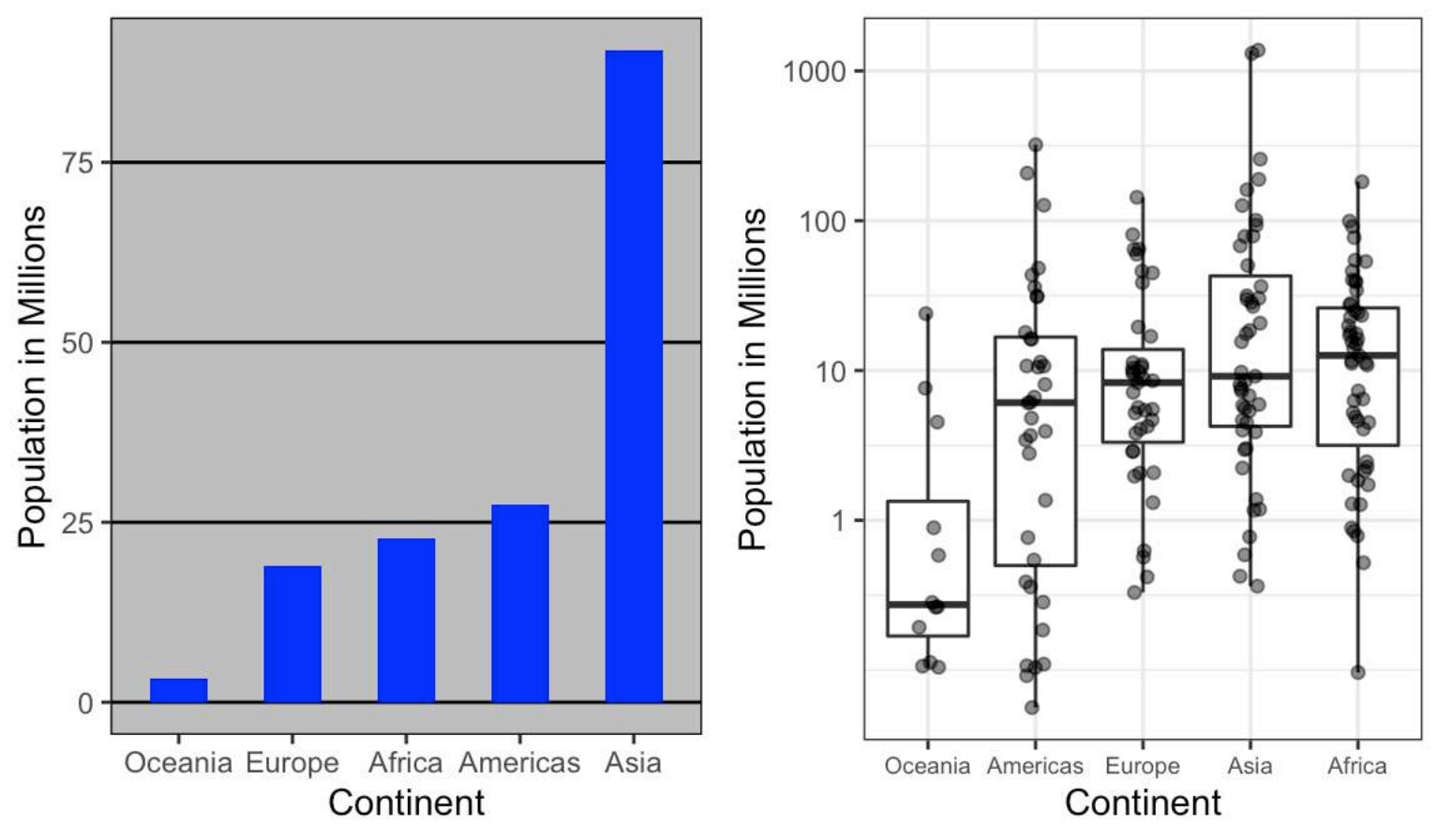
Principles of Data Visualization: **4. Ease comparisons**

Principles of Data Visualization: **4. Ease comparisons**

- **(iv) Consider transformations**

Principles of Data Visualization: **4. Ease comparisons**

- **(iv) Consider transformations**



Principles of Data Visualization: **5. Think of the color blind**

Principles of Data Visualization: **5. Think of the color blind**

- **Seaborn color pallettes**
- Color Brewer website provides some guidance on **which palettes are color blind safe.**
- Seaborn choose_colorbrewer_palette: interactive widget to browse options and tweak parameters; must be used in a Jupyter notebook
- Example code to understand how the the seaborn color palettes compare for different type of colorblindness (e.g., deuteranopia, protan)
- There is a variety of kinds of color blindness, but the **most common** variant leads to **difficulty distinguishing reds and greens**.
- Seaborn colorblind palette:

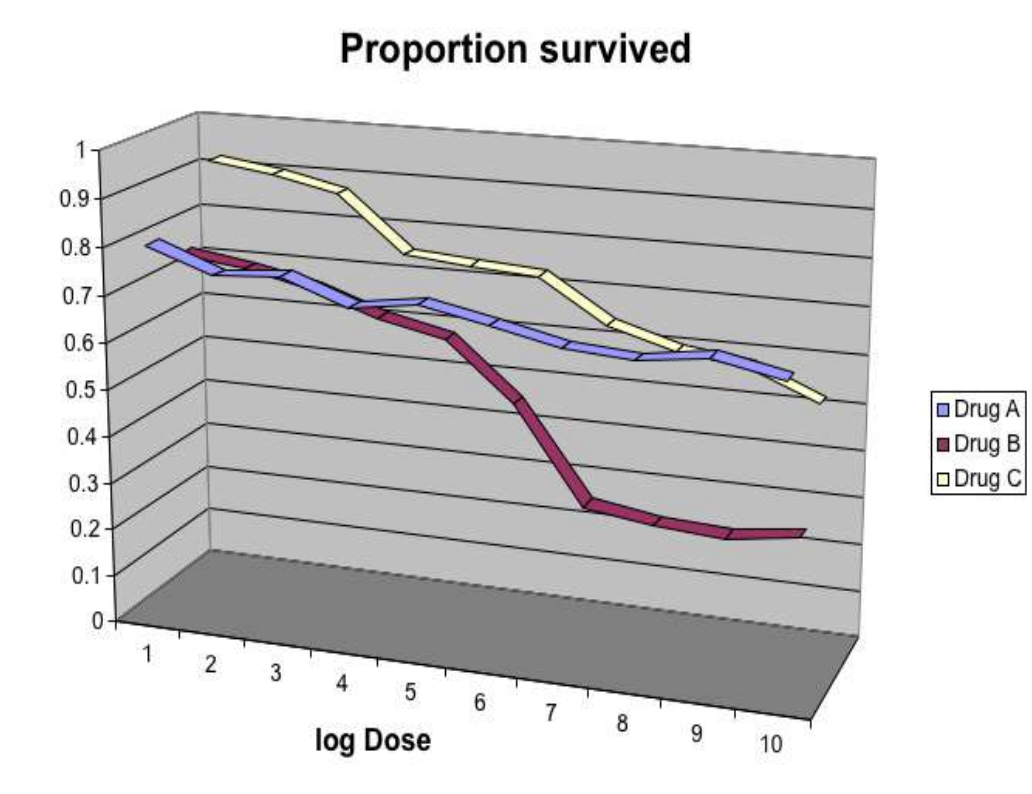
Principles of Data Visualization: **5. Think of the color blind**

- **Seaborn color pallettes**
- Color Brewer website provides some guidance on **which palettes are color blind safe.**
- Seaborn choose_colorbrewer_palette: interactive widget to browse options and tweak parameters; must be used in a Jupyter notebook
- Example code to understand how the the seaborn color palettes compare for different type of colorblindness (e.g., deuteranopia, protan)
- There is a variety of kinds of color blindness, but the **most common** variant leads to **difficulty distinguishing reds and greens**.
- Seaborn colorblind palette:

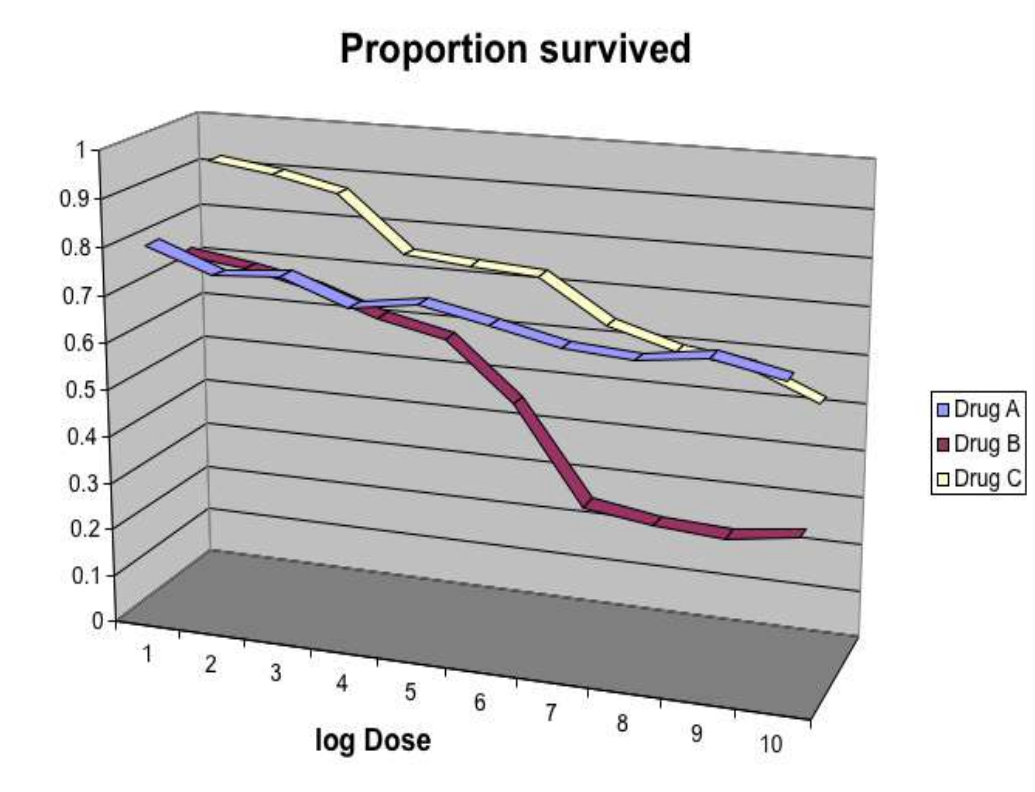
*Example code: color_pal = sns.color_palette("**colorblind**", 6).as_hex()*

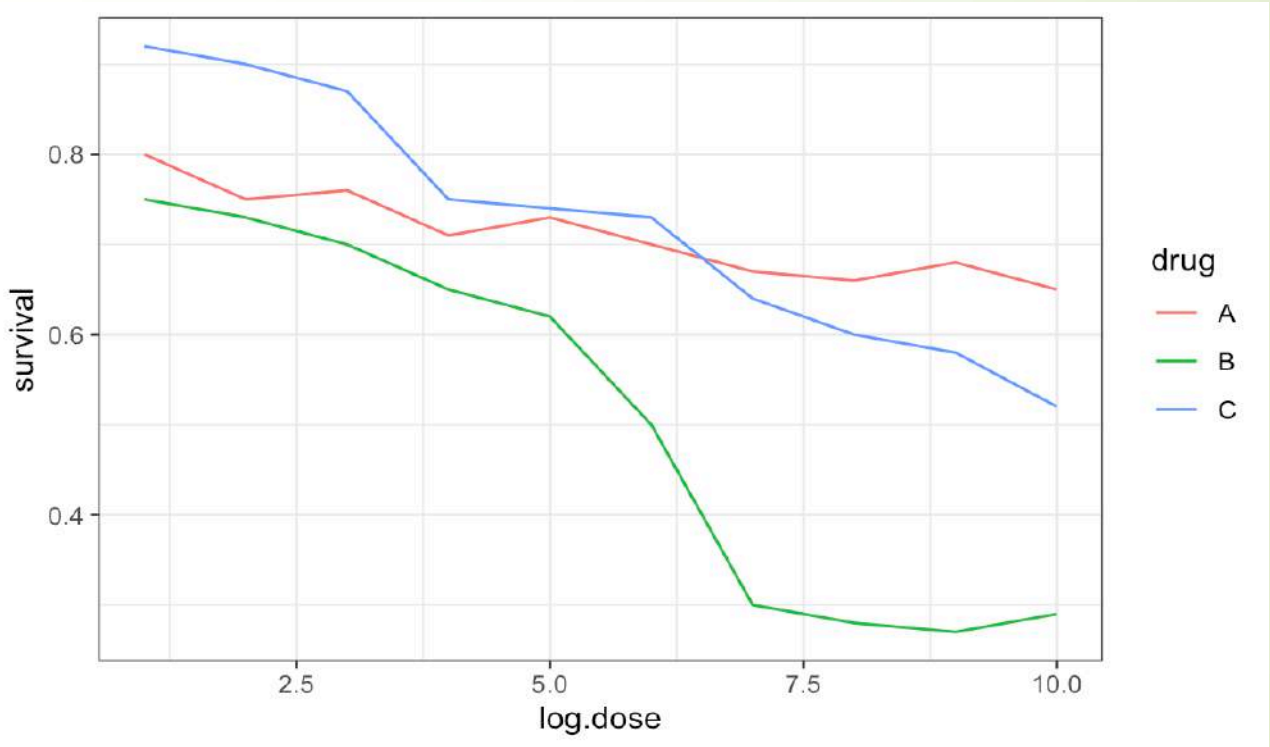
Principles of Data Visualization: **6. Avoid pseudo-three-dimensional plots**

Principles of Data Visualization: **6. Avoid pseudo-three-dimensional plots**



Principles of Data Visualization: **6. Avoid pseudo-three-dimensional plots**





Principles of Data Visualization: **7. Avoid too many significant digits**

Principles of Data Visualization: **7. Avoid too many significant digits**

state	year	Measles	Pertussis	Polio
California	1940	37.8826320	18.3397861	0.8266512
California	1950	13.9124205	4.7467350	1.9742639
California	1960	14.1386471	NA	0.2640419
California	1970	0.9767889	NA	NA
California	1980	0.3743467	0.0515466	NA

Principles of Data Visualization: **7. Avoid too many significant digits**

state	year	Measles	Pertussis	Polio
California	1940	37.8826320	18.3397861	0.8266512
California	1950	13.9124205	4.7467350	1.9742639
California	1960	14.1386471	NA	0.2640419
California	1970	0.9767889	NA	NA
California	1980	0.3743467	0.0515466	NA

state	year	Measles	Pertussis	Polio
California	1940	37.9	18.3	0.8
California	1950	13.9	4.7	2.0
California	1960	14.1	NA	0.3
California	1970	1.0	NA	NA
California	1980	0.4	0.1	NA

Principles of Data Visualization: **8. Place values being compared on columns rather than on rows**

Principles of Data Visualization: **8. Place values being compared on columns rather than on rows**

state	year	Measles	Pertussis	Polio
California	1940	37.9	18.3	0.8
California	1950	13.9	4.7	2.0
California	1960	14.1	NA	0.3
California	1970	1.0	NA	NA
California	1980	0.4	0.1	NA

Principles of Data Visualization: **8. Place values being compared on columns rather than on rows**

state	year	Measles	Pertussis	Polio
California	1940	37.9	18.3	0.8
California	1950	13.9	4.7	2.0
California	1960	14.1	NA	0.3
California	1970	1.0	NA	NA
California	1980	0.4	0.1	NA

state	disease	1940	1950	1960	1970	1980
California	Measles	37.9	13.9	14.1	1	0.4
California	Pertussis	18.3	4.7	NA	NA	0.1
California	Polio	0.8	2.0	0.3	NA	NA