Data visualization

Yue Jiang

Duke University

A disclaimer

The following material was used during a live lecture. Without the accompanying oral comments and discussion, the text is incomplete as a record of the presentation. A full recording may be found via Zoom on the course Sakai site.

Why visualization?

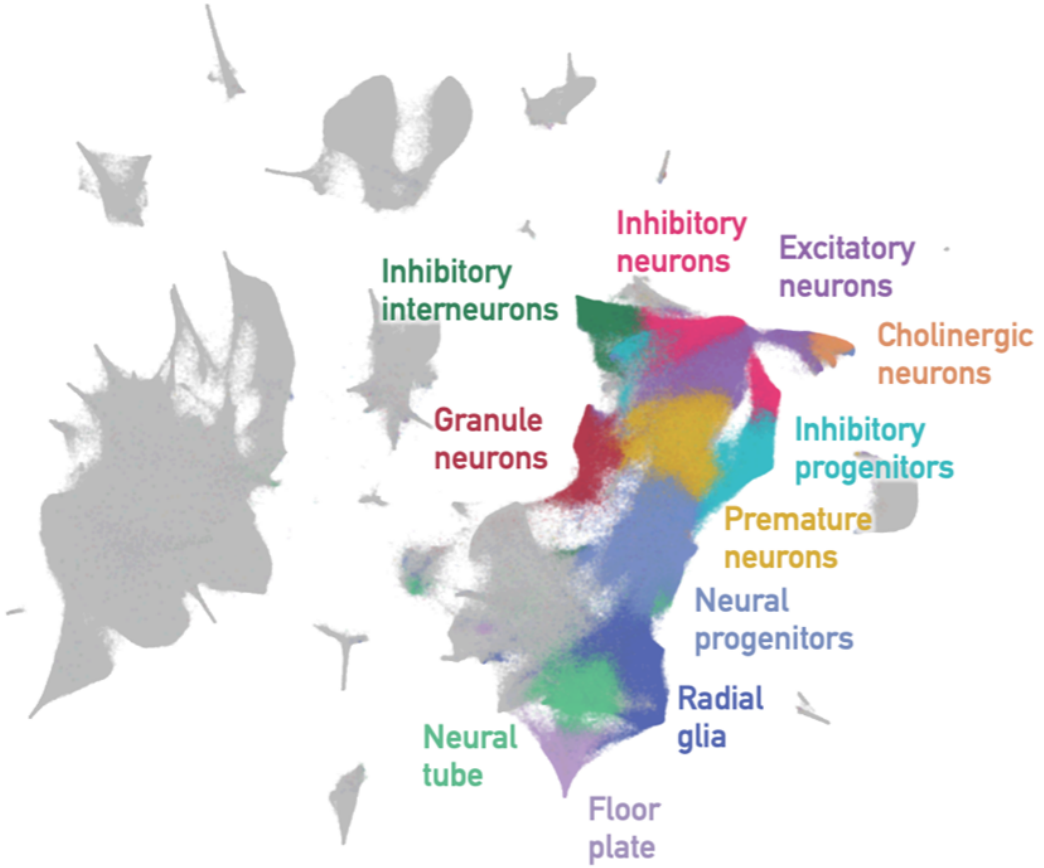


Image source: Adapted from Wang, 2020, re: Kobak and Berens (Nat. Commun. 2019) and Cao et al. (Nature 2019).

Why visualization?

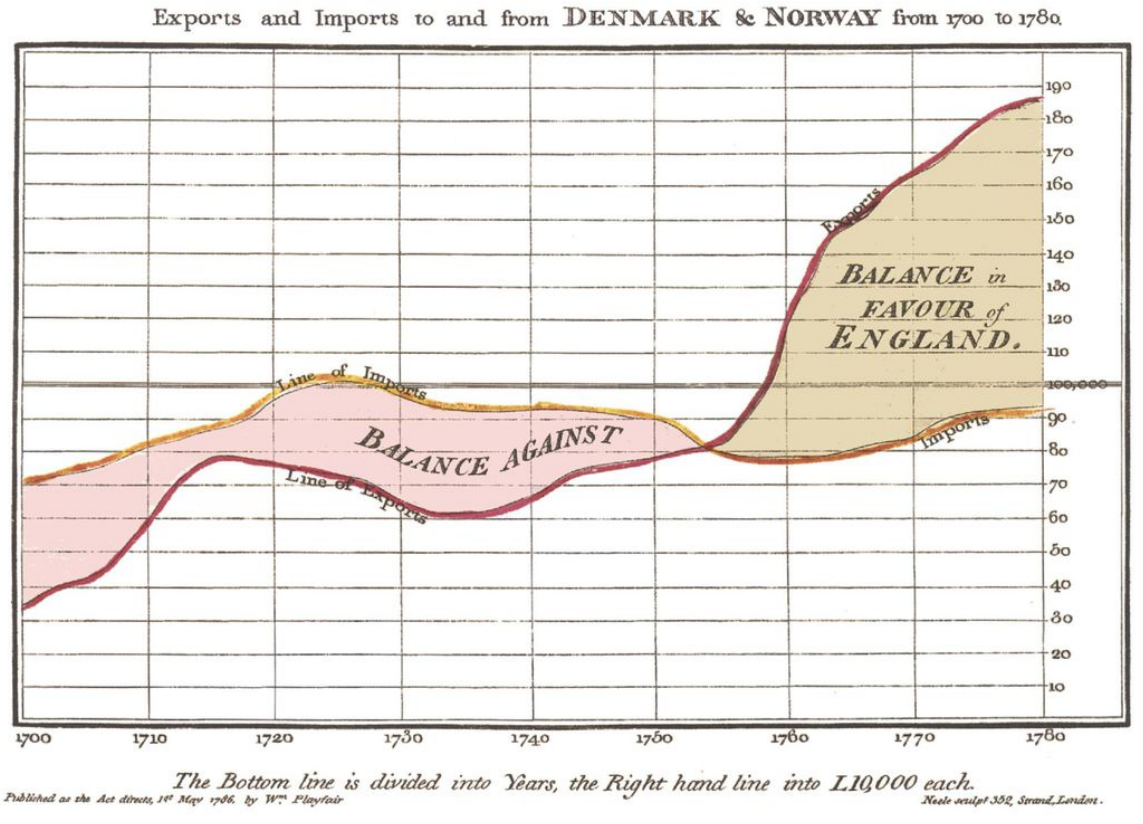


Image source: Playfair. 1786 trade balance chart, Commercial and Political Atlas.

Why visualization?

```
dat
```

```
##    x1 x2 x3 x4    y1   y2    y3    y4
## 1  10 10 10  8  8.04 9.14  7.46  6.58
## 2   8  8  8  8  6.95 8.14  6.77  5.76
## 3  13 13 13  8  7.58 8.74 12.74  7.71
## 4   9  9  9  8  8.81 8.77  7.11  8.84
## 5  11 11 11  8  8.33 9.26  7.81  8.47
## 6  14 14 14  8  9.96 8.10  8.84  7.04
## 7   6  6  6  8  7.24 6.13  6.08  5.25
## 8   4  4  4 19  4.26 3.10  5.39 12.50
## 9  12 12 12  8 10.84 9.13  8.15  5.56
## 10  7  7  7  8  4.82 7.26  6.42  7.91
## 11  5  5  5  8  5.68 4.74  5.73  6.89
```

Why visualization?

```
dat %>%
  summarize(meanx1 = mean(x1),
            meanx2 = mean(x2),
            meanx3 = mean(x3),
            meanx4 = mean(x4))
```

```
##   meanx1 meanx2 meanx3 meanx4
## 1      9      9      9      9
```

```
dat %>%
  summarize(meany1 = mean(y1),
            meany2 = mean(y2),
            meany3 = mean(y3),
            meany4 = mean(y4))
```

```
##   meany1 meany2 meany3 meany4
## 1    7.5    7.5    7.5    7.5
```

Why visualization?

```
dat %>%
  summarize(sdx1 = var(x1),
            sdx2 = var(x2),
            sdx3 = var(x3),
            sdx4 = var(x4))
```

```
##   sdx1 sdx2 sdx3 sdx4
## 1   11   11   11   11
```

```
dat %>%
  summarize(sdy1 = sd(y1),
            sdy2 = sd(y2),
            sdy3 = sd(y3),
            sdy4 = sd(y4))
```

```
##   sdy1 sdy2 sdy3 sdy4
## 1 2.03 2.03 2.03 2.03
```

Why visualization?

```
dat %>%
  summarize(cor1 = cor(x1, y1),
            cor2 = cor(x2, y2),
            cor3 = cor(x3, y3),
            cor4 = cor(x4, y4))
```

```
##    cor1  cor2  cor3  cor4
## 1 0.816 0.816 0.816 0.817
```

Why visualization?

```
lm(dat$y1 ~ dat$x1)$coef
```

```
## (Intercept)      dat$x1
##         3.0         0.5
```

```
lm(dat$y2 ~ dat$x2)$coef
```

```
## (Intercept)      dat$x2
##         3.0         0.5
```

```
lm(dat$y3 ~ dat$x3)$coef
```

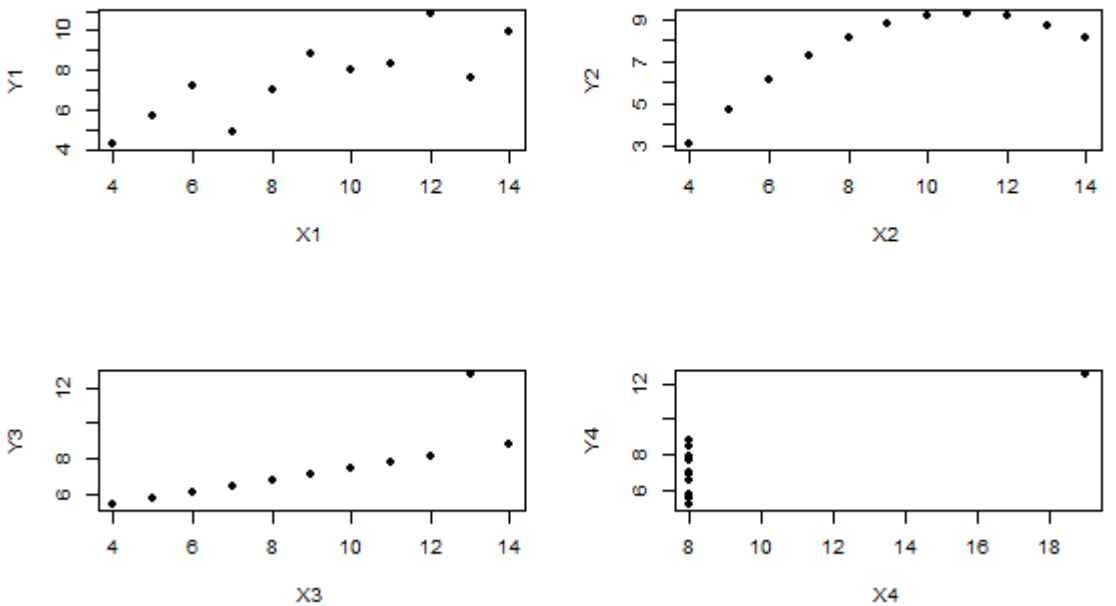
```
## (Intercept)      dat$x3
##         3.0         0.5
```

```
lm(dat$y4 ~ dat$x4)$coef
```

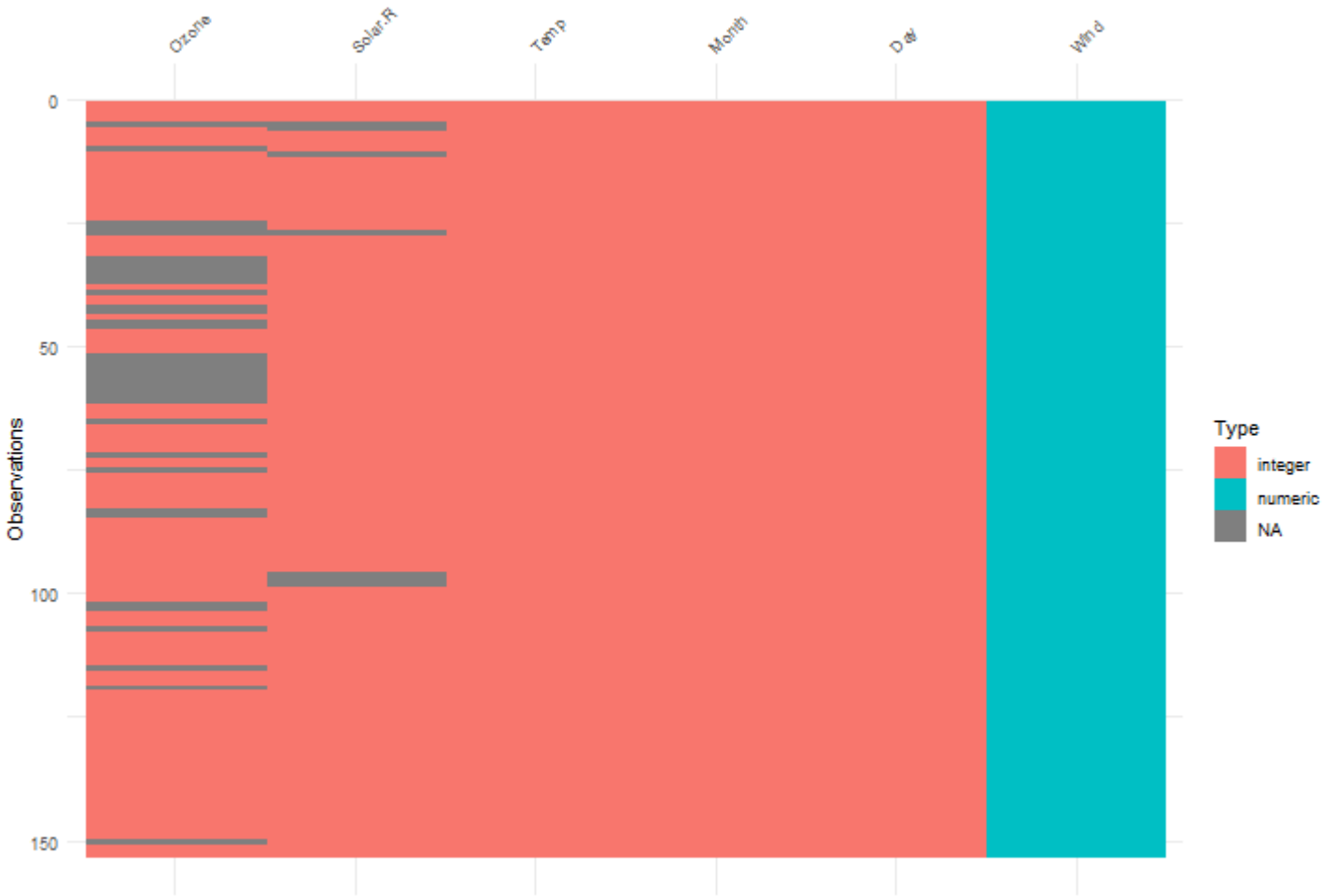
```
## (Intercept)      dat$x4
##         3.0         0.5
```

Why visualization?

```
par(mfrow = c(2,2))
plot(dat$x1, dat$y1, xlab = "X1", ylab = "Y1", pch = 19)
plot(dat$x2, dat$y2, xlab = "X2", ylab = "Y2", pch = 19)
plot(dat$x3, dat$y3, xlab = "X3", ylab = "Y3", pch = 19)
plot(dat$x4, dat$y4, xlab = "X4", ylab = "Y4", pch = 19)
```



Why visualization?



What can go wrong?

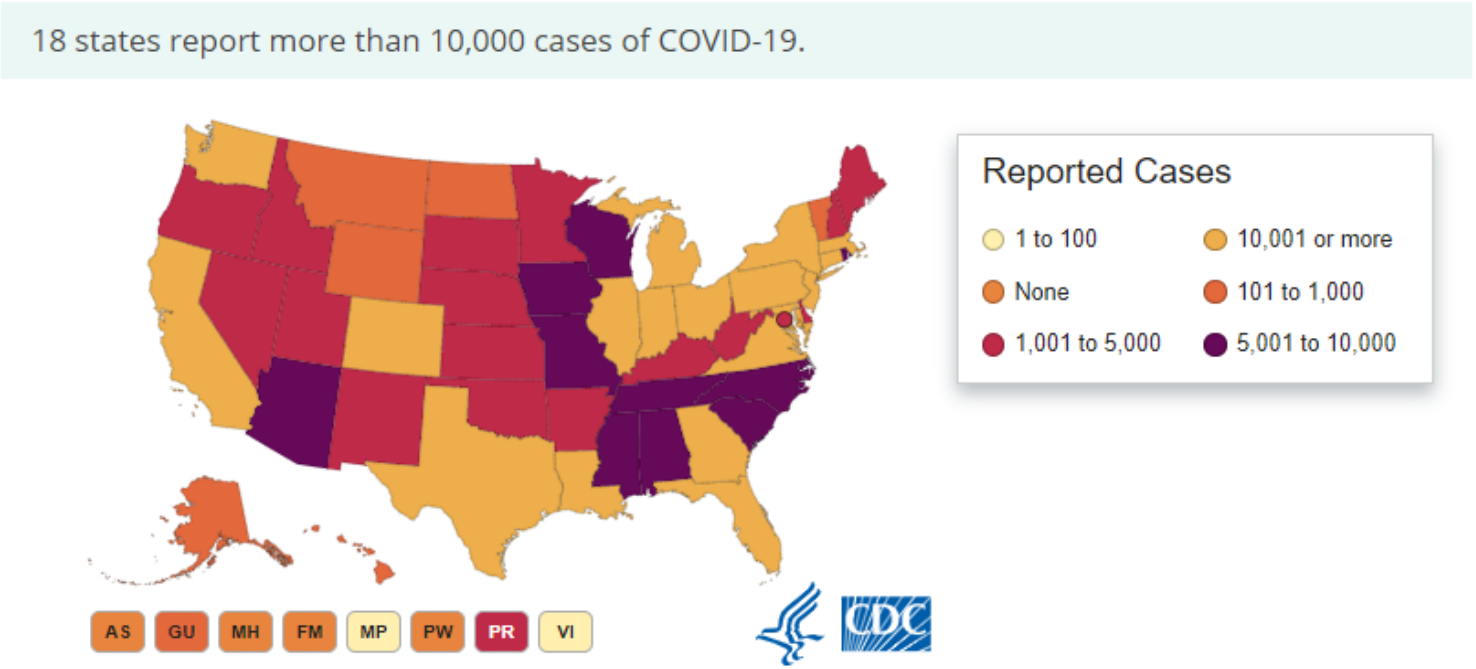


Image source: CDC, 2020

What can go wrong?

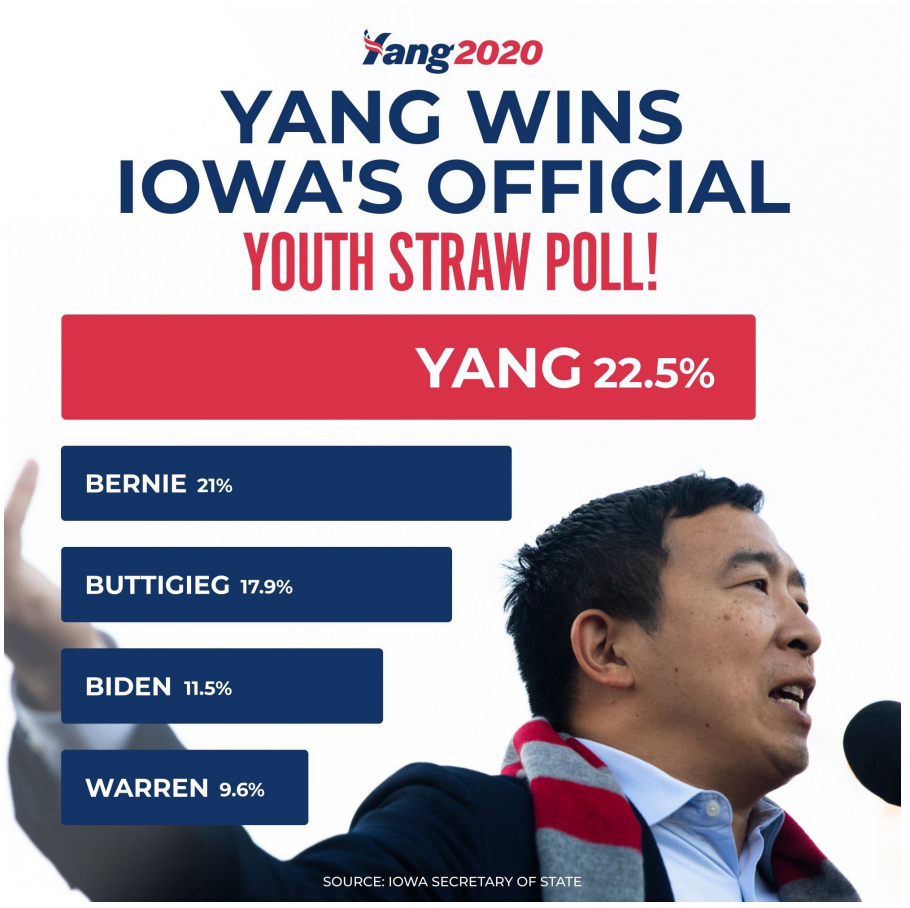
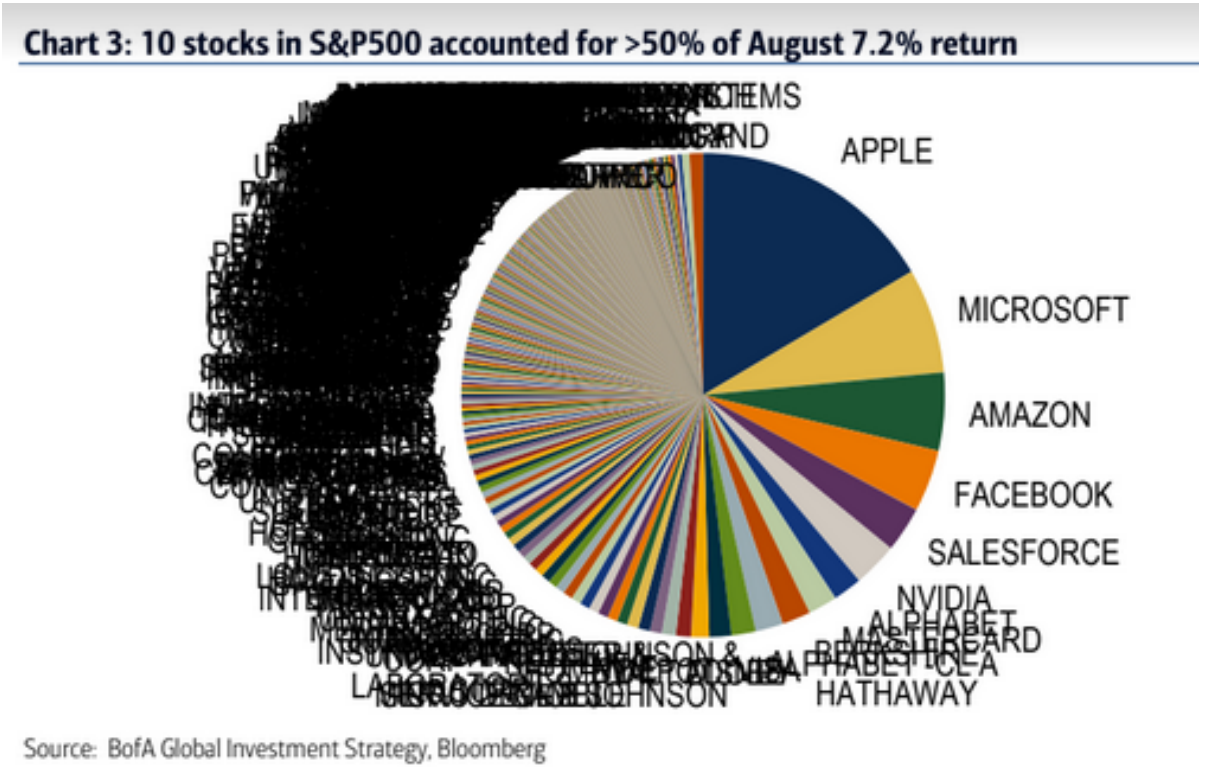
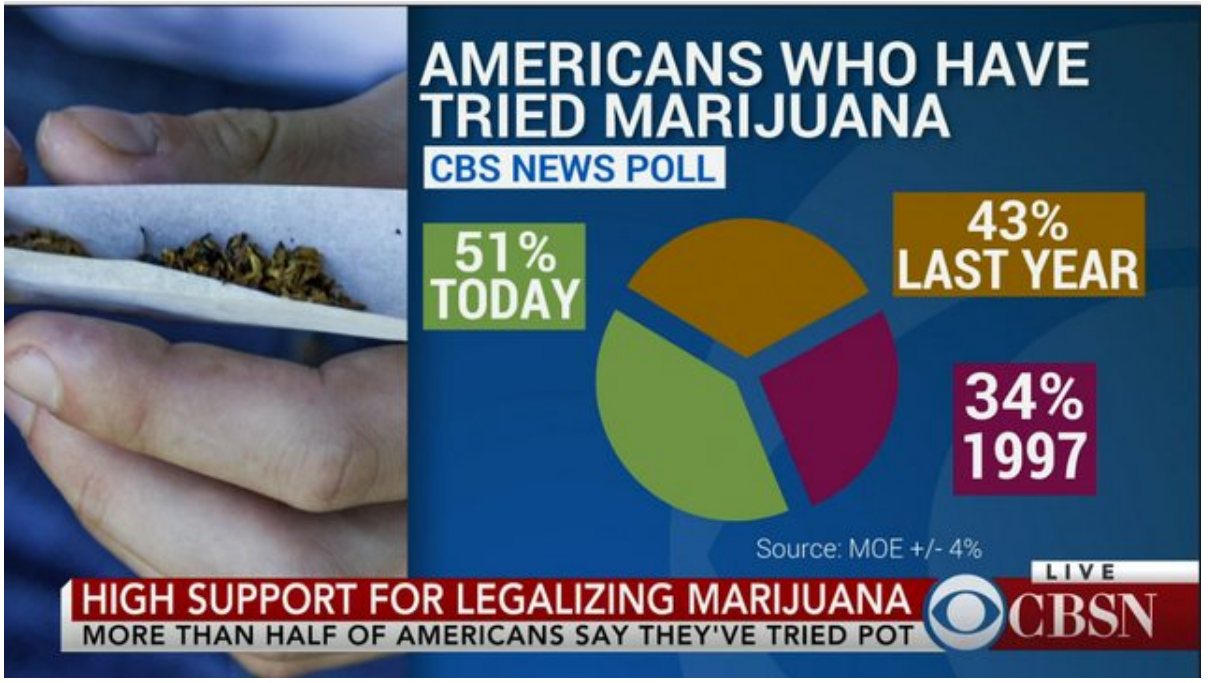


Image source: Andrew Yang campaign, 2020

What can go wrong?



What can go wrong?



What can go wrong?

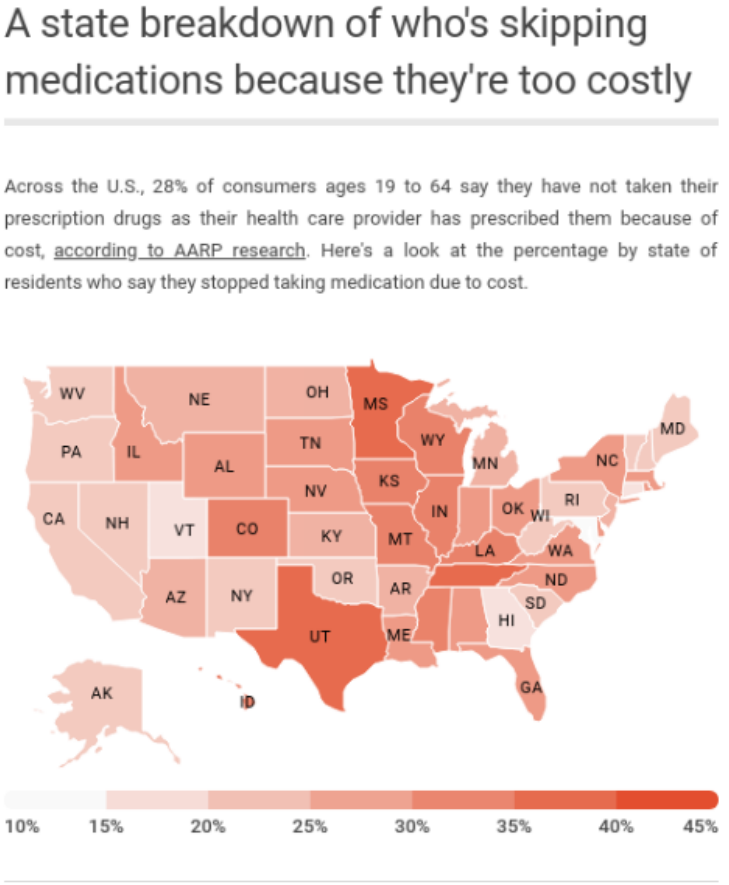
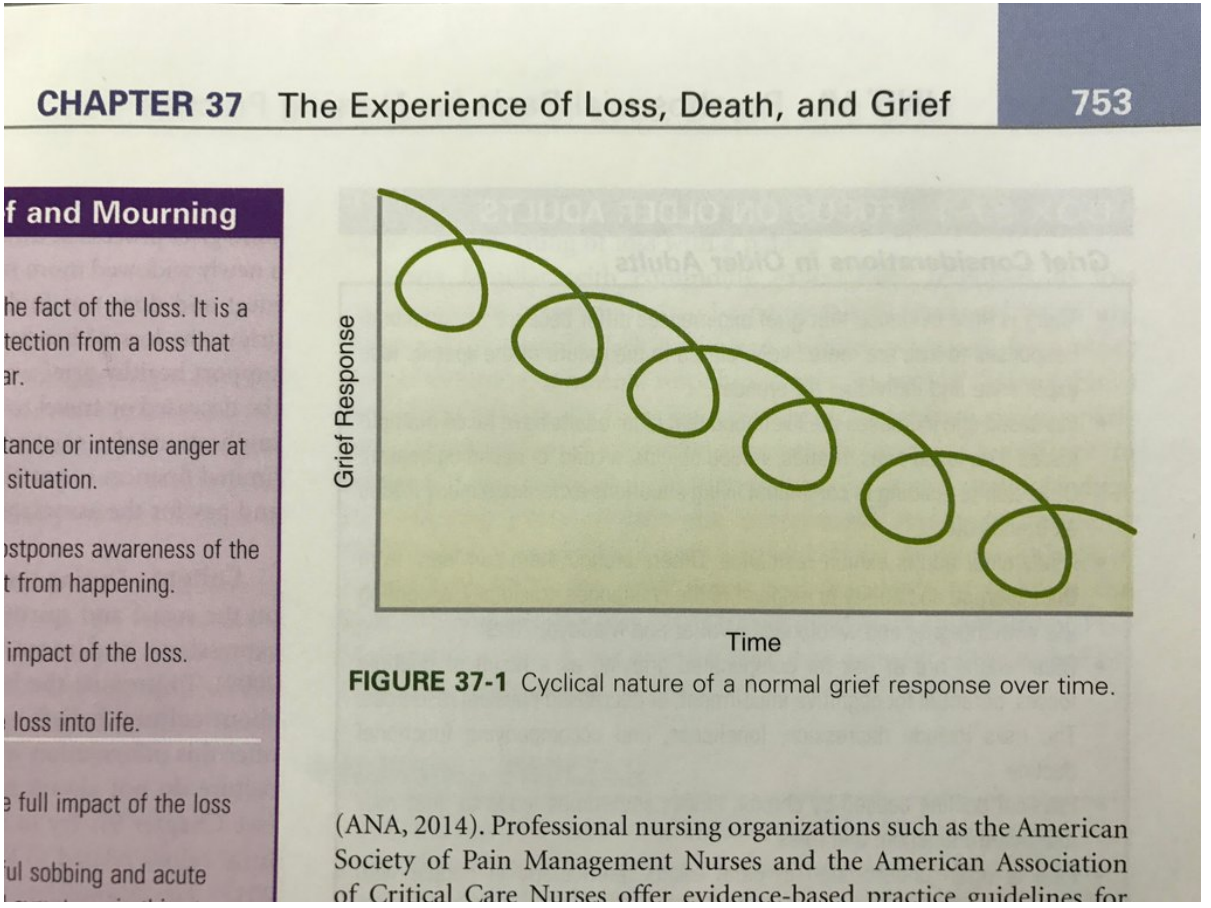


Image source: Leonhardt 2020, CNBC

What can go wrong?



CHAPTER 37 The Experience of Loss, Death, and Grief 753

f and Mourning

he fact of the loss. It is a
tection from a loss that
ar.

tance or intense anger at
situation.

stpones awareness of the
t from happening.

impact of the loss.

loss into life.

e full impact of the loss

ul sobbing and acute

Grief Response

Time

FIGURE 37-1 Cyclical nature of a normal grief response over time.

(ANA, 2014). Professional nursing organizations such as the American Society of Pain Management Nurses and the American Association of Critical Care Nurses offer evidence-based practice guidelines for

Image source: Potter et al., Fundamentals of Nursing

What can go wrong?

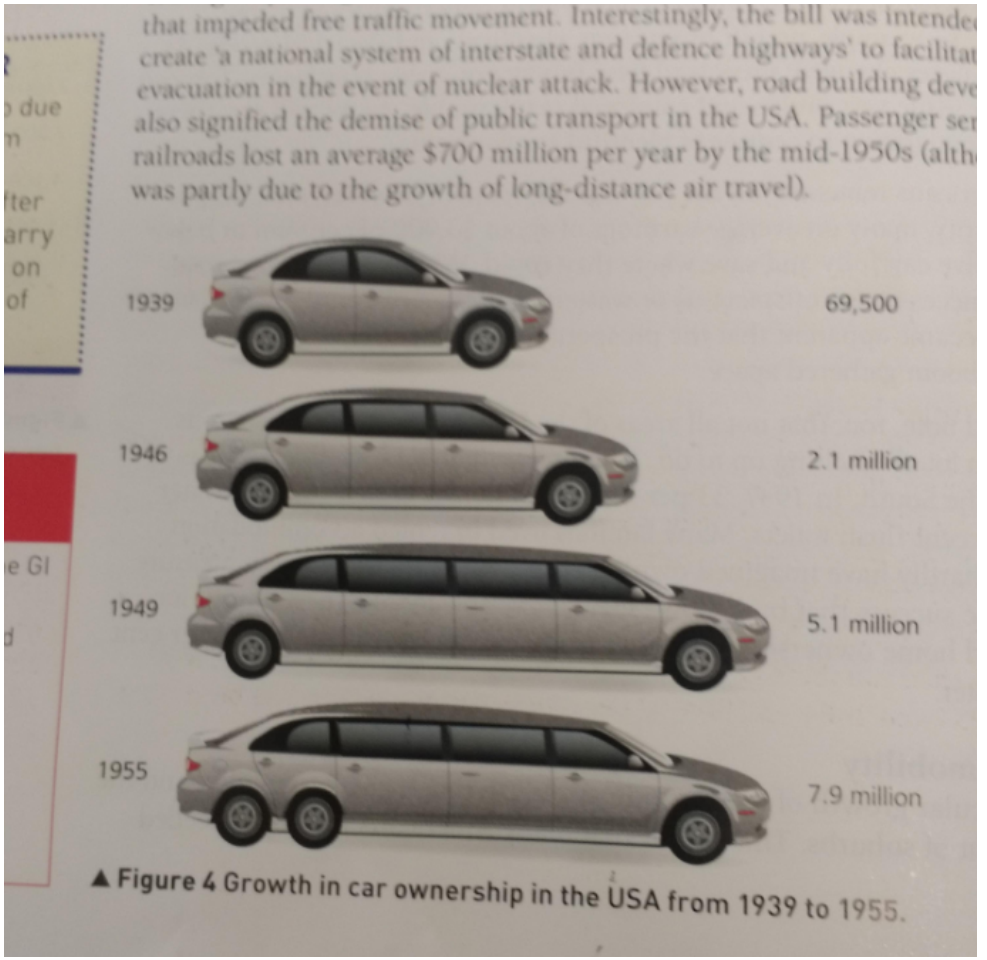


Image source: Waugh and Clements, AQA-level History: Making of a Superpower: US 1865-1975

What can go wrong?

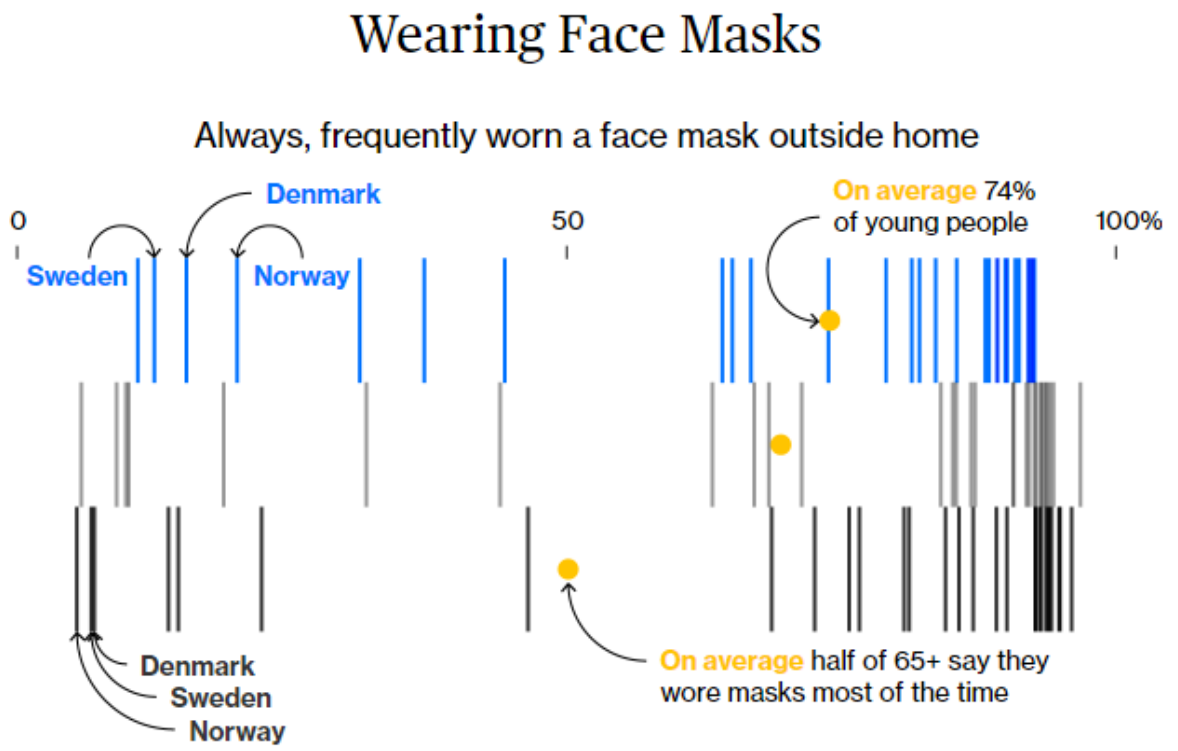


Image source: He and Williams for Bloomberg, 2020

What can go wrong?

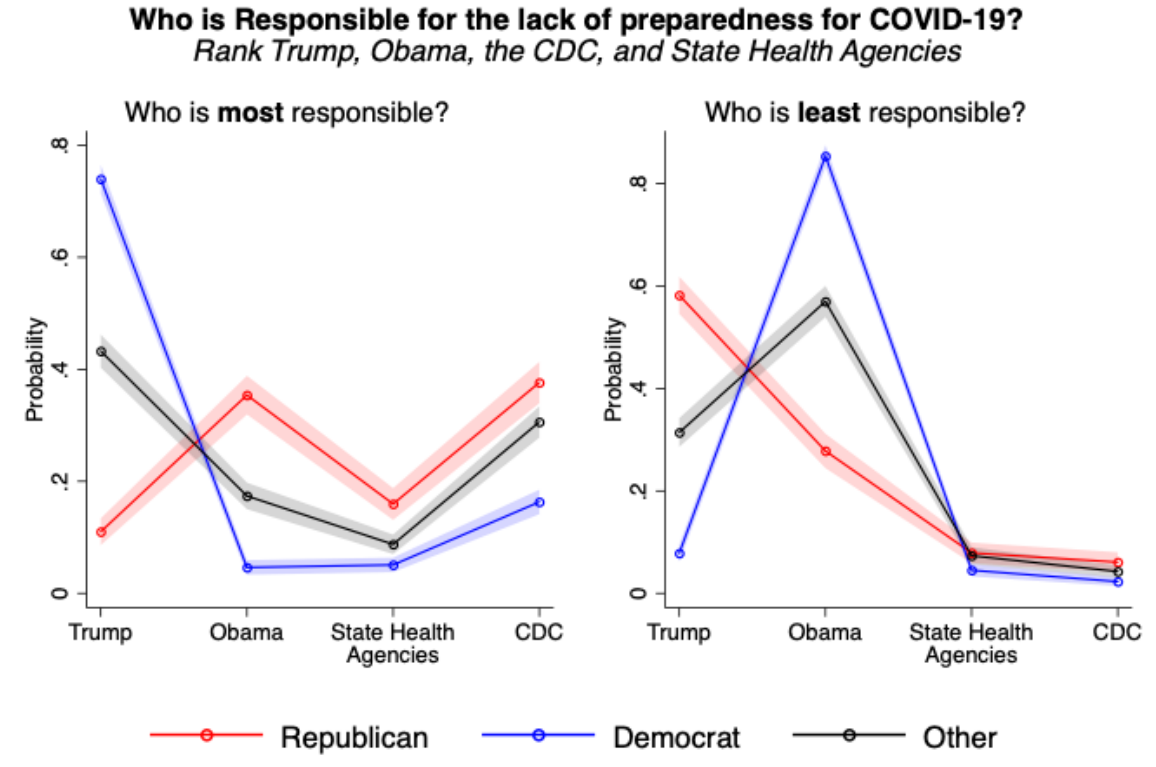


Image source: Gadarian, Goodman, and Pepinsky (Draft figure from SSRN 2020)

What can go wrong?

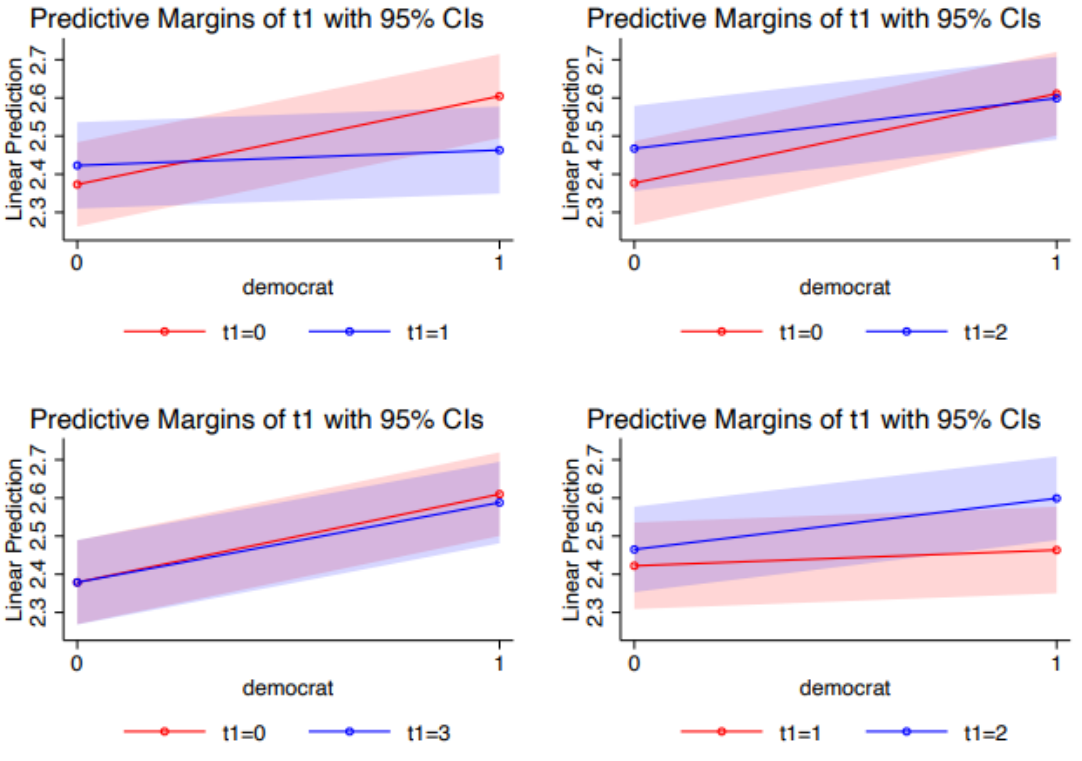


Image source: Gadarian, Goodman, and Pepinsky (SSRN 2020)

Some best practices

- Tell a story; have one in mind when deciding to use a visualization
- Every figure should serve a clear purpose
- A figure should stand on its own
- Keep it simple and clear

Some best practices

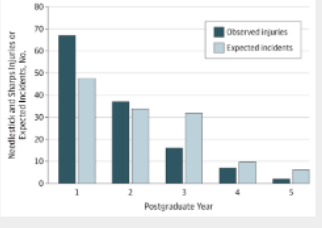
Make sure the type of visualization chosen is correct given the number of variables included, their types, and their relationships. Some helpful links are below:

- The Data Visualization Catalogue
- The Financial Times Visual Vocabulary

Some best practices

Often, journals will present their own guidelines for visualizations. For instance, to submit an article to JAMA, a style guide for figures and their purposes is provided here and contains the most often used plots for this particular journal.

Table of Figure Requirements

Figure Type	Correct Usage and Creation
Bar graph 	To present frequency data (numbers or percentages). Each bar represents a category. Bar graphs are typically vertical but when categories have long titles or there are many of them, they may run horizontally. The scale on the frequency axis should begin at 0, and the axis should not be broken. If the data plotted are a percentage or rate, error bars may be used to show statistical variability. **Acceptable File Formats for Initial Submission:** .ai, .bmp, .docx, .emf, .eps, .jpg, .pdf, .ppt, .psd, .tif, .wmf, .xls **Acceptable File Formats for Revision and Publication:** .ai, .emf, .eps, .pdf, .wmf, .xls

Workshop