1 Statistical Graphics

No unique answer. Here is an example.

(a) In Figure 1, the horizontal bars clutter the graph and don’t bring much information. We don’t need them to see any trend.

(b) In Figure 2, I would like the axis labels to read simply "total" and "date" without the data set name "harare", and I would prefer that the y-axis scale annotation (50, 100, etc.) be horizontal, not vertical.

Something weird: the dip in Fig 1 does not occur in Fig 2. I think it has nothing to do with the graphing software, but simply that the dipping point was not present to make fig.2, due to some bug. Not the same data were used.

(c) In Figure 3, I like that the points are not connected: no need to connect the variation in # of new cases from week to week (would make a jagged pattern, too cluttered). I like the various comments and annotation on the graph itself: we don’t need to read the legend to get the main points. If there’s a way to add arrows and comments without add too much cluttering, I would add an arrow pointing to the second bump and add a comment "unexpected second bump".

2 Academic Article

No unique answer. Here is an example.


(b) The sample consists of 194 healthy male students. They were from different universities in Zurich. Exclusion criteria were significant medical or psychiatric illness, medication, smoking more than 15 cigarettes per day, and drug or alcohol abuse.

(c) The biological basis for trust in humans.

(d) Intranasal administration of oxytocin, a neuropeptide that plays a key role in social attachment and affiliation in non-human mammal, causes a substantial increase in trust among humans, thereby greatly increasing the benefits from social interactions. They also show that the effect of oxytocin on trust is not due to a general increase in the readiness to bear risks. On the contrary, oxytocin specifically affects an individual’s willingness to accept social risks arising through interpersonal interactions. These results concur with animal research suggesting an essential role for oxytocin as a biological basis of prosocial approach behaviour.
(e) The average transfer amount per investor in Trust Experiment and Risk Experiment is displayed. The horizontal axis represents average transfer amount per investor, in MU (monetary units). The vertical axis represents relative frequency. Also, oxytocin group and placebo group are displayed by filled bars and open bars, respectively. This graph shows that oxytocin increases investors considerably in the trust experiment and that investors' behaviour does not differ between the oxytocin group and placebo group.

(f) Average transfer amount and oxytocin are two of the variables used in this article. The means for oxintocin group and placebo group are reported, 9.6 MU and 8.1 MU, respectively. In this study, average transfer amount is numerical, observational, and response, while oxytocin use is categorical, experimental, and explanatory.

3 Problem 13 in Chapter 1

(a) Observational. Because the two groups are naturally assigned instead of being assigned by the researchers.

(b) Subspecies and sensitivity of retina.


4 Chapter 2

4.1 Problem 16

(a) Frequency table.
(b) One.

(c) 21.

(d) $265/395 = 0.6709$.

(e) A line graph would be appropriate, since it allows us to see the distribution of number of convictions and the variable is discrete. A bar plot or a histogram would also suit our purpose.
R commands for line graph:

```r
nconv <- seq(0, 14, 1)
counts <- c(265, 49, 21, 19, 10, 10, 2, 2, 4, 2, 1, 4, 3, 1, 2)
plot(nconv, counts, type = "h",
     main = "Frequency Distribution of Number of Convictions",
     xlab = "Number of Convictions",
     ylab = "Frequency")
```

R commands for bar plot:

```r
counts <- c(265, 49, 21, 19, 10, 10, 2, 2, 4, 2, 1, 4, 3, 1, 2)
barplot(counts, beside = TRUE, names.arg = c(0:14),
        main = "Frequency Distribution of Number of Convictions",
        xlab = "Number of Convictions")
```

R commands for histogram:

```r
breaks <- seq(0, 15, 1)
counts <- c(265, 49, 21, 19, 10, 10, 2, 2, 4, 2, 1, 4, 3, 1, 2)
conv <- list(breaks = breaks, counts = counts)
class(conv) <- "histogram"
plot(conv, main = "Frequency Distribution of Number of Convictions",
     xlab = "Number of Convictions")
```

(f) It is skewed and unimodal with mode 0. No outliers.

(g) No. The boys are chosen from 6 schools that are near the research office.
4.2 Problem 17

(a) Contingency table.

(b) Mosaic plot.

R commands:

```r
conv.inc <- matrix(c(47, 43, 128, 57, 90, 30), nrow = 2, ncol = 3)
colnames(conv.inc) <- c("Inadequate", "Adequate", "Comfortable")
rownames(conv.inc) <- c("No", "Yes")
mosaicplot(conv.inc, col = c("black", "grey", "white"), dir = c("v", "h"), main = "Relation Between Income Level and Convction")
```
R commands:

```r
mosaicplot(t(conv.inc), col = c("black","grey","white"), dir = c("v", "h"),
main = "Relation Between Income Level and Convction")
```

(c) Categorical, ordinal. And in this case, income level serves as explanatory variable, so the second display is better.

(d) As the income level increase, say from Inadequate to Comfortable, the proportion of conviction decreases.

(e) The graph. It is more informative than the table, showing the proportion of each income level and the proportion of conviction within each level.

4.3 Problem 26

(a) Contingency table.

<table>
<thead>
<tr>
<th></th>
<th>0 sneaking male</th>
<th>1</th>
<th>2 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannibalized</td>
<td>61</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>No Cannibalization</td>
<td>389</td>
<td>17</td>
<td>4</td>
</tr>
</tbody>
</table>

(b) Mosaic plot. R commands:

```r
can <- matrix(c(61, 389, 18, 17, 16, 4), nrow = 2, ncol = 3)
colnames(can) <- c("0 sneaking male", "1", "2 or more")
rownames(can) <- c("Cannibalized", "No Cannibalization")
mosaicplot(t(can), col = c("black", "white"), dir=c("v", "h"),
main = "Cannibalization")
```
5 Chapter 3

5.1 Problem 10

(a) 5.5.

(b) 0.2555.

(c) 100% of the numbers fall within two standard deviation of the mean, [4.99, 6.01].

5.2 Problem 14

(a) Histogram.

(b) 1000, the histogram is approximately symmetric.

(c) 1000, same reason as in (b).

(d) 1000, the histogram is approximately symmetric and the interval [1000,1100) has the highest frequency.

(e) 200. By the Empirical Rule, there should be approximately 68%, or 2200 observations falling within one standard deviation of the mean. [800, 1200] contains about 2250 observations, hence can be treated as the one-standard-deviation range.