

BDSI R Markdown: Exercise 2

type your name here

14-June-2021

Contents

Main header	1
Sub header	1
Keeping track of your margins	1
Some data analysis	1
The Datasaurus dozen	2

Main header

This exercise assumes you are reading the R script and not the knitted output. **Task 1** Add yourself as the author of this document on line 3.

Task 2 *Knit* this document using *Cmd+Shift+K* / *Ctrl+Shift+K*. Note how Markdown creates an automatic table of contents with click-able links.

Task 3 On line 7 change `html_document` to `word_document`, and on line 11 change `word_document` to `html_document`. **knitr** will pay attention only to the first output selection and ignore all others. Don't worry if what you knit doesn't look very pretty as a .docx. Now undo what you just did so that your selected output format is again a html

Sub header

Note the clever use of the `#` sign at the beginning of each line number. As usual, “regular” R will ignore any line that starts with a `#`, as these do. Thus, you can treat this script exactly as you would any other R script. But, when you knit this script using **knitr**, the `#'` (the hash followed by an apostrophe) indicates to that this is actually markdown text and should be interpreted as such. **Task 4** Add some more text after this sentence and then press *enter* or *return*. RStudio should automatically added an `#'` at the beginning of each new line that you start.

Keeping track of your margins

Your knitted document is not affected by single linebreaks. That is, you can press ‘enter’ once without affecting the appearance of your final document. On the other hand, two linebreaks denote a new paragraph. Take advantage of this feature to maintain the readability of your text by starting a new line before you start to spill off the screen of your text editor, like this: Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

On my RStudio application, I have set it to denote an 80-character width, which is the approximate point at which I try to jump to a new line when I'm typing.

Task 6 To do this on your own app: go to the RStudio menu, then Preferences > Code > Display > Show Margin. Then enter your desired width. After pressing ‘Ok’, you should see a slight vertical line somewhere towards the right of your screen (depending on your chosen number).

Some data analysis

I'm using an in-line chunk to report that the observed mean of `x` is -1.9811 **Task 7** Add another in-line chunk to report the observed standard deviation of `x`.

The Datasaurus dozen

```
if(!require(datasauRus, quietly = TRUE)) {install.packages("datasauRus",repos = "http://cran.us.r-project.org")

# If the package is installed and you just want to load it, you can also use this code:

library(datasauRus);
```

The first chunk you ran above checks to see if a particular package (`datasauRus`, keep reading for a description of these packages) is *installed* on your computer, i.e. downloaded onto your hard drive. After installing a package, you generally only ever need to re-install it if you've updated R or if the package is updated. Thus, the chunk first checks to see if each package has already been installed before trying to re-install it.

Here's what the top of the `datasaurus` dataset looks like in raw form:

```
as_tibble(datasaurus_dozen);
```

```
## # A tibble: 1,846 x 3
##   dataset      x      y
##   <chr>   <dbl> <dbl>
## 1 dino    55.4  97.2
## 2 dino    51.5  96.0
## 3 dino    46.2  94.5
## 4 dino    42.8  91.4
## 5 dino    40.8  88.3
## 6 dino    38.7  84.9
## 7 dino    35.6  79.9
## 8 dino    33.1  77.6
## 9 dino    29.0  74.5
## 10 dino   26.2  71.4
## # ... with 1,836 more rows
```

There are 13 individual datasets included in this dataset, and we can summarize the characteristics of each dataset with this code:

```
datasaurus_summarized <-
  datasaurus_dozen %>%
  group_by(dataset) %>%
  summarize(mean_x = mean(x),
            mean_y = mean(y),
            std_dev_x = sd(x),
            std_dev_y = sd(y),
            corr_x_y = cor(x, y));
print(datasaurus_summarized)
```

```
## # A tibble: 13 x 6
##   dataset    mean_x mean_y std_dev_x std_dev_y corr_x_y
##   <chr>      <dbl> <dbl>    <dbl>    <dbl>    <dbl>
## 1 away      54.3  47.8     16.8     26.9  -0.0641
## 2 bullseye  54.3  47.8     16.8     26.9  -0.0686
## 3 circle    54.3  47.8     16.8     26.9  -0.0683
## 4 dino      54.3  47.8     16.8     26.9  -0.0645
## 5 dots      54.3  47.8     16.8     26.9  -0.0603
## 6 h_lines   54.3  47.8     16.8     26.9  -0.0617
## 7 high_lines 54.3  47.8     16.8     26.9  -0.0685
## 8 slant_down 54.3  47.8     16.8     26.9  -0.0690
## 9 slant_up   54.3  47.8     16.8     26.9  -0.0686
## 10 star     54.3  47.8     16.8     26.9  -0.0630
## 11 v_lines   54.3  47.8     16.8     26.9  -0.0694
## 12 wide_lines 54.3  47.8     16.8     26.9  -0.0666
## 13 x_shape   54.3  47.8     16.8     26.9  -0.0656
```

Here is how these same summary data look like using **knitr**'s **kable** command:

```
kable(datasaurus_summarized);
```

dataset	mean_x	mean_y	std_dev_x	std_dev_y	corr_x_y
away	54.27	47.83	16.77	26.94	-0.0641
bullseye	54.27	47.83	16.77	26.94	-0.0686
circle	54.27	47.84	16.76	26.93	-0.0683
dino	54.26	47.83	16.77	26.94	-0.0645
dots	54.26	47.84	16.77	26.93	-0.0603
h_lines	54.26	47.83	16.77	26.94	-0.0617
high_lines	54.27	47.84	16.77	26.94	-0.0685
slant_down	54.27	47.84	16.77	26.94	-0.0690
slant_up	54.27	47.83	16.77	26.94	-0.0686
star	54.27	47.84	16.77	26.93	-0.0630
v_lines	54.27	47.84	16.77	26.94	-0.0694
wide_lines	54.27	47.83	16.77	26.94	-0.0666
x_shape	54.26	47.84	16.77	26.93	-0.0656

We can also use the **kableExtra** package to make our tables extra fancy. However, this makes use of some features that will not work for all file formats.

Task 8 To see this, go back to the YAML header of this document and, again, On line 7 change `html_document` to `word_document`, and on line 11 change `word_document` to `html_document`. After knitting, go back to this table: it's just a jumbled mess. Do the same thing, now knitting to a pdf (knitting to a pdf will only work if you have a Tex distribution installed). The knitted pdf should have stripes; the html version will not.

The takeaway is this: although you *can* always knit your script to different formats, it will not always be formatted as expected. MS Word files are very different from html, which are very different from pdf files. I find pdf files to be the most pleasing to look at and easiest to share with collaborators.

```
if(!require(kableExtra, quietly = TRUE)) {install.packages("kableExtra",repos = "http://cran.us.r-project.org")}

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##   group_rows

kable(datasaurus_summarized,
      booktabs = T) %>%
  kable_styling(latex_options = c("striped"),
               stripe_color = "#DFDFDF") %>%
  add_header_above(c(" " = 1, "Summary statistics" = 5));
```

Visually, the datasets look much different:

```
dino_plot <-
  ggplot(datasaurus_dozen) +
  geom_point(mapping = aes(x = x, y = y, color = dataset)) +
  theme_grey() +
  theme(legend.position = "none") +
  facet_wrap(~dataset, ncol = 3);
print(dino_plot);
```

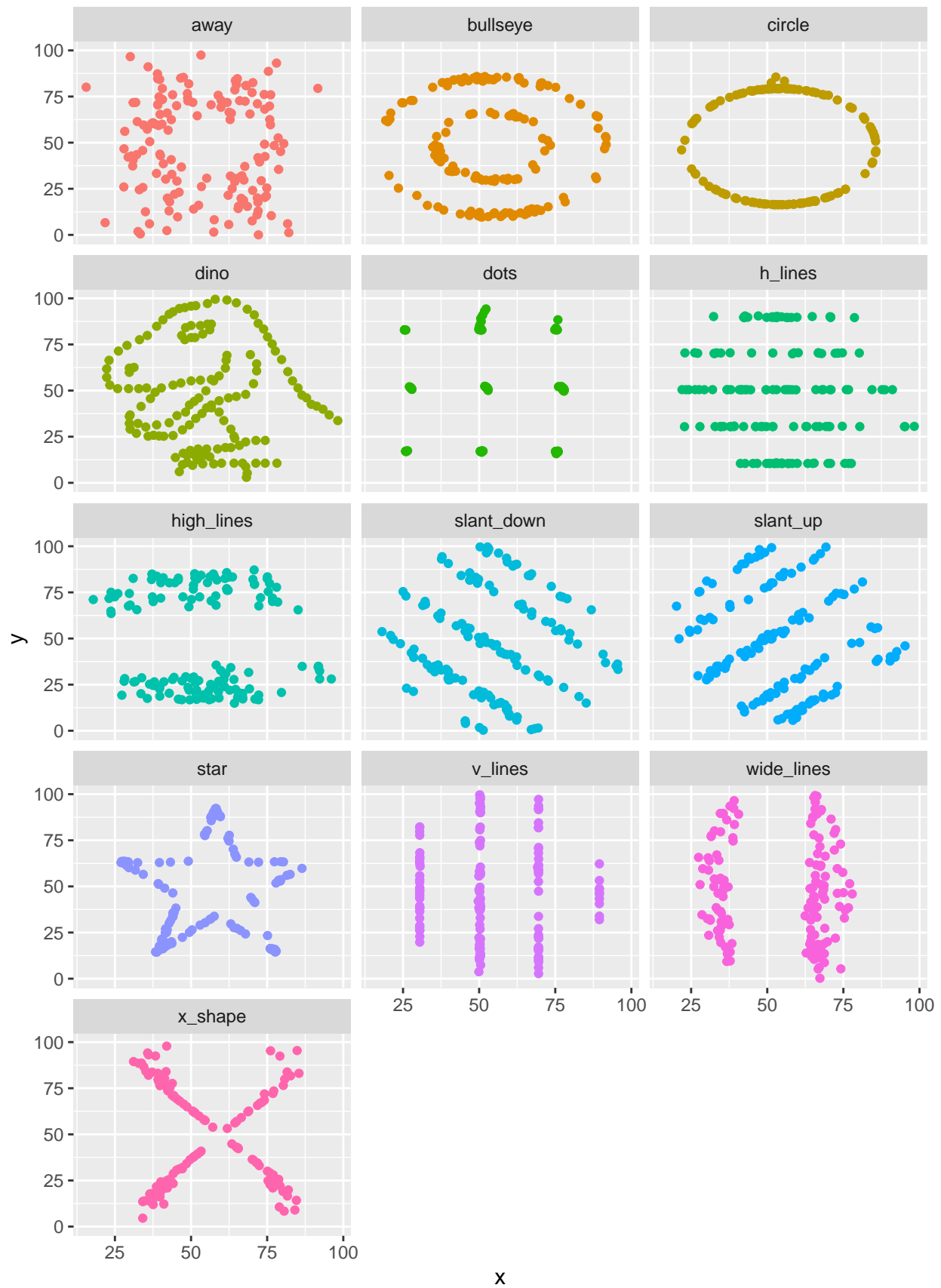


Figure 1: Same summary statistics, different plots. Roar!

Summary statistics					
dataset	mean_x	mean_y	std_dev_x	std_dev_y	corr_x_y
away	54.27	47.83	16.77	26.94	-0.0641
bullseye	54.27	47.83	16.77	26.94	-0.0686
circle	54.27	47.84	16.76	26.93	-0.0683
dino	54.26	47.83	16.77	26.94	-0.0645
dots	54.26	47.84	16.77	26.93	-0.0603
h_lines	54.26	47.83	16.77	26.94	-0.0617
high_lines	54.27	47.84	16.77	26.94	-0.0685
slant_down	54.27	47.84	16.77	26.94	-0.0690
slant_up	54.27	47.83	16.77	26.94	-0.0686
star	54.27	47.84	16.77	26.93	-0.0630
v_lines	54.27	47.84	16.77	26.94	-0.0694
wide_lines	54.27	47.83	16.77	26.94	-0.0666
x_shape	54.26	47.84	16.77	26.93	-0.0656