**Collaborative notes for Plotting in R Workshop on December 1, 2021**
**Workshop webpage: https://microbiaki.github.io/2021-12-01-WHOI-plotting/**
**Tutorial lessons: https://microbiaki.github.io/workshop_t2/**

**Answers to chat questions and comments:**

13:18:38 From Jessica Drysdale to Everyone:
I have used R limited and always just used the console (on Mac) - how to do I get a view like Maria with the various quadrants?
13:18:57 From (Instructor) Sarah Hu (she/her) to Everyone:
It is via RStudio
13:19:30 From (Instructor) Sarah Hu (she/her) to Everyone:
Requires an install of RStudio - and it should be able to find R that you already have installed. Instructions on the website.
13:19:47 From (Helper) Stace Beaulieu to Everyone:
Link to download and install RStudio is at bottom of workshop schedule webpage
13:20:44 From (Helper) Brett Longworth (he/him) to Everyone:
RStudio is highly recommended, but I think all of what we do today will work in the console. Let us know if you don't see a plot from your console code when Maria finishes the first plot.

13:42:20 From Jessica Drysdale to Everyone:
I am seeing plots as I think it should look but keep getting the message “Warning message:
Removed 11 rows containing missing values (geom_point)” - is this oK?

- Warning messages such as: “Removed 11 rows containing missing values (geom_point)” are totally ok. It is just a warning about missing values.
  13:43:42 From (Helper) Cory Berger (he/him) to Everyone:
  R is just letting you know that there are some missing values in your data, but it will plot everything else

13:47:40 From Sarah Glancy | she/her to Everyone:
why is shape outside aesthetics?
13:48:42 From (Helper) Brett Longworth (he/him) to Everyone:
@Sarah, Shape is being set as a constant, rather than data mapping to shape
13:49:25 From (Helper) Brett Longworth (he/him) to Everyone:
Think: aes() maps data to what you see, constant values set what you see to the constant.

13:53:40 From Laura W Stolp to Everyone:
What is the library needed for the stat_regline?
13:53:50 From (Helper) Brett Longworth (he/him) to Everyone:
ggpubr

- ggpubr library is needed to plot the trend line equation

14:01:08 From (Helper) Brett Longworth (he/him) to Everyone:
The "sex ~ species" notation is called formula notation. You'll also see it in modelling functions like "lm()"
14:10:54 From Kali Horn (she/her) to Everyone:
I love having the multiple variables in facet ! I have only ever used 1 at a time that is super useful 🙂
14:11:54 From (Helper) Brett Longworth (he/him) to Everyone:
facet_grid() is where it's at!

- **Challenge answer:**
  ggplot(penguins, aes(x=flipper_length_mm, y=body_mass_g, fill=species))+
  + geom_point(shape=24)

14:14:25 From Kate Lane to Everyone:
What does position= identity mean?
- position="identity" tells the histogram to plot the groups overlapping another. Default position (position="stack"), will create stacked bars.
- 14:17:52 From (Helper) Cory Berger (he/him) to Everyone:
- @kate, position="identitiy" tells the histogram to plot the groups overlapping another. If you use position="stack", you the bars would be stacked on top of one another on the y axis (might be easiest to see if you try it out yourself)
- 14:19:14 From (Helper) Cory Berger (he/him) to Everyone:
- position='stack' is the default for bar plots, but it's often not what you want
- 14:19:37 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
- And the default (if you do not specify position) will be ”stack”.
- 14:20:03 From (Helper) Brett Longworth (he/him) to Everyone:
- also, position = "stack" is the default. The position parameter controls how a geom is shifted when it's plotted. "identity" means "leave it where the data tell it to be". We'll see another example of using position to shift points to keep them from overlapping later...
-

14:14:47 From Jessica Drysdale to Everyone:
how do you get the GGplot theme gallery?
14:15:36 From (Helper) Stace Beaulieu to Everyone:
@Jessica you can click on theme options here https://r-charts.com/ggplot2/themes/
- Sarah’s favorite theme: theme_linedraw()

14:21:54 From Wanyi Lu to Everyone:
Is there a way to add x axis intervals? For example, label 180, 190, 200, 210, 220?
14:22:56 From (Helper) Brett Longworth (he/him) to Everyone:
@Wanyi, yes, this is controlled by the "scales" functions

- You can think of the "scale" series of functions as creating a space to map data onto with a geom. So here, we're creating a group of colors that data can be mapped onto.
- scale_x_continuous(breaks=seq(170, 230, 10))#change ticks
- 14:23:53 From (Helper) Brett Longworth (he/him) to Everyone:
- @wanyi - https://ggplot2.tidyverse.org/reference/scale_continuous.html
-
- 14:52:09 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
- scale_x_continuous(breaks=seq(180, 220,10))
- 14:53:14 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
- Or scale_x_continuous(breaks=seq(160, 240,10))?
- 14:55:44 From (Helper) Brett Longworth (he/him) to Everyone:
- The scales package has a bunch of functions for generating nice axis scale breaks. scale_x_continuous(breaks = scales::breaks_extended(n=5)) and play with n.
-

14:25:21 From (Helper) Brett Longworth (he/him) to Everyone:
Also the scales chapter in the ggplot book is great!
https://ggplot2-book.org/scale-position.html

- How do you get the bars to sit on the x-axis and not hovering above it?
  + scale_y_continuous(expand = c(0,0))

14:26:46 From (Instructor) Sarah Hu (she/her) to Everyone:
Challenge: Remake this plot, remove the legend and the x-axis labels.
14:29:27 From Olivia Ahern to Everyone:

```
ggplot(penguins, aes(flipper_length_mm, fill=species)) +
  geom_histogram(alpha=0.8, position="identity")+
  theme_linedraw() +
  labs(x="Flipper Length (mm)", y = "Counts") +
  theme(axis.title.x=element_text(color = "black", face="bold",size=14),
        axis.title.y=element_text(color = "black", face="bold",size=14),
        axis.text.x = element_blank(),
        axis.text.y = element_text(color='gray30',size=12),
        # modify legend
        legend.title=element_blank(),
        legend.position = "none")
```

15:07:35 From (Helper) Brett Longworth (he/him) to Everyone:
Note that just like geom_histogram, the default for geom_bar is position = "stack"

- **To specify input data for ggplot you have 2 options:**
  **# Option 1:**
  **ggplot(penguins, aes(...**
  15:12:25 From (Helper) Brett Longworth (he/him) to Everyone:

The first argument/parameter to ggplot() is always the data you want to plot

**# Option 2:**
**penguins %>%**
**ggplot(aes(...**

15:11:50 From Lyndsey Lefebvre (she/her) to Everyone:
How do you get the bars to sit on the x-axis and not hovering above it?
15:12:23 From (Instructor) Sarah Hu (she/her) to Everyone:
Try this: scale_y_continuous(expand = c(0,0))
15:23:26 From (Helper) Stace Beaulieu to Everyone:
Hi, this is answer for @Lyndsey a while back in chat: ggplot defaults to giving you a little space around the extent of the axis labels. scale_y_continuous(expand = c(0,0)) will set this scale expansion to zero

15:15:53 From Jessica Drysdale to Everyone:
very rudimentary question but how can I open new plot window to be able to keep an old one to compare? I have previously saved them but assume this is the slow way…
15:16:48 From (Instructor) Sarah Hu (she/her) to Everyone:
good question, the arrows in the left top corner can be used to go back and forth!
15:19:03 From (Helper) Brett Longworth (he/him) to Everyone:
@Jessica, I don't think the back and forward arrows for plots are available in the basic R GUI. RStudio gives you this function

15:20:16 From Kate Lane to Everyone:
I often manually click "Run" in RStudio, is there a keycode to run a line of code in R studio?
- Run shortcuts: (Mac) Command + Enter, (Windows) ctrl+shift

15:21:19 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
All the shortcuts are at the end of the tutorial (Further reading)

- **Remove outliers from box plot:** outlier.shape=NA

15:39:40 From (Helper) Brett Longworth (he/him) to Everyone:
position_dodge() requires a grouping variable, position_dodge2() does not. dodge() uses fill = sex in ggplot(aes()) to know how we want to dodge the boxplots.

- **Challenge add labels**: patchwork + plot_annotation(tag_levels = ‘A’)

- 15:44:10 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
- Challenge: add labels (A and B) in your panels
- 15:44:21 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
- Or “a” and “b”
- 15:45:51 From Kali Horn (she/her) to Everyone:
- For patchwork + plot_annotation(tag_levels = ‘A’)
- 15:46:02 From (Helper) Stace Beaulieu to Everyone:
- https://patchwork.data-imaginist.com/articles/patchwork.html#annotating-the-composition-1
- 15:46:57 From Kali Horn (she/her) to Everyone:
- You can also use #s
- 15:49:40 From (Helper) Brett Longworth (he/him) to Everyone:
- One cool thing about patchwork is it extends the "+" idea for combining plots: horizontal_bar / violin_mod will stack them vertically.
- 15:50:51 From Kali Horn (she/her) to Everyone:
- What if you want 2 different y ?

15:51:46 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
https://www.r-graph-gallery.com/line-chart-dual-Y-axis-ggplot2.html
15:53:18 From Kate Lane (she/her) to Everyone:
Would love resources to learn more about vector graphics!

- I will add in the tutorial how to export svg with ggsave

16:08:58 From Wanyi Lu to Everyone:
what is the difference between using ggsave and pushing the buttons of "Export" under "Plots"?
16:13:44 From Sarah Glancy | she/her to Everyone:
is this just saving the most recent plot sicne it is not specified in the code?
16:13:56 From Jessica Drysdale to Everyone:
should my plot now be saved after ggsave? Not getting an error but not seeing it in the file
16:14:18 From (Instructor) Sarah Hu (she/her) to Everyone:
ggsave() will save the last plot.
try running "getwd()" to find out what directory ggsave is saving in
16:16:16 From Kali (she/her) to Everyone:
You can specify which plot you want in ggsave though if you name it
16:17:26 From (Helper) Brett Longworth (he/him) to Everyone:
@wanyi, if you use Rmd, you can also grab saved plots from the output. Setting dev="svg" in the chunk options will render as svg
16:15:26 From (Helper) Brett Longworth (he/him) to Everyone:
most web browsers can view svg

**Tips to adjust tick labels:**

- https://stackoverflow.com/questions/7263849/what-do-hjust-and-vjust-do-when-making-a-plot-using-ggplot
- http://www.sthda.com/english/wiki/ggplot2-axis-ticks-a-guide-to-customize-tick-marks-and-labels
- https://ggplot2-book.org/scale-position.html

**Tips to choose colors:**

- https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3
- viridis is designed to improve graph readability for readers with common forms of color blindness and/or color vision deficiency https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
- 14:39:05 From (Helper) Brett Longworth (he/him) to Everyone:
- You can think of the "scale" series of functions as creating a space to map data onto with a geom. So here, we're creating a group of colors that data can be mapped onto.
-

**Tip: great package for 'skimming' data tables: skimr**
**https://cran.r-project.org/web/packages/skimr/vignettes/skimr.html**

**Question about packages for plotting maps in R:**

- tutorial for Spatial Data in R developed by JP alum Chrissy Hernandez: [link to Google slides]
- Data Carpentry has lessons Introduction to R for Geospatial Data https://datacarpentry.org/r-intro-geospatial/ and Introduction to Geospatial Raster and Vector Data with R https://datacarpentry.org/r-raster-vector-geospatial/
- **Let us know if you'd like a future workshop on mapping with R (+1 in Chat)**

16:22:32 From (Instructor) Maria Pachiadaki (she/her) to Everyone:
https://oceanview.pfeg.noaa.gov/erddap
16:24:18 From (Helper) Stace Beaulieu to Everyone:
For more about - what is ERDDAP - see Karen Soenen's materials from last month workshop: https://k-rns.github.io/workshop_data_reuse/01_open_data_erddap/index.html
16:25:25 From (Helper) Brett Longworth (he/him) to Everyone:
ERDDAP has good support in R through the rerddap package as well, even though it really should have been named erddapr.

**If you are exploring the data and do not want to download the file, you can pull it directly into your R environment**

- This allows you to avoid downloading the file:
  d <- read.csv("https://oceanview.pfeg.noaa.gov/erddap/tabledap/cciea_AC.csvp")

16:31:40 From (Helper) Stace Beaulieu to Everyone:
I always grab the csvp from erddap so as to avoid the additional row at top (which you get with csv)

16:39:26 From Sarah Glancy | she/her to Everyone:
lubridate has been really helpful for me switching time formats
16:40:50 From (Helper) Stace Beaulieu to Everyone:
@Sarah I agree lubridate good package to work with datetime

16:40:10 From Kali (she/her) to Everyone:
What is sub("-.+", "" doing exactly?
16:40:34 From (Instructor) Sarah Hu (she/her) to Everyone:
It is using 'regular expression' to do a substitution. So…
anything after the "-" will be removed and turned into "" which is nothing!
16:41:20 From (Instructor) Sarah Hu (she/her) to Everyone:
The ".+" after the dash is a wild card reg expression syntax to select anything
16:41:50 From (Helper) Brett Longworth (he/him) to Everyone:
The regular expression approach to getting the time data is powerful, but converting to a datetime class is less error-prone
16:42:05 From (Helper) Stace Beaulieu to Everyone:
For more about regular expressions see Amber York's lesson materials https://adyork.github.io/regex-intro/01-regex-intro/index.html

Additional help:

- Slack whoi-datascience
- Brett mentioned Tidy Tuesday
- https://www.rfordatasci.com/
- The ggplot book