Data Visualization

Choosing Visualization

Scatter Plots visualize the relationship between 2 quantitative variables.
Line Plots compares non-numeric to quantitative data.
Bar Plots visualize comparisons of amounts.
Histograms visualize the distribution of one quantitative variable.

Geometric Objects

```
geom_point(alpha=1)
geom_line()
geom_histogram(alpha  = ..., position = “identity”, binwidth=...)
geom_bar(stat = “identity”)

geom_vline(xintercept = ..., linetype = “”, size = ... )
geom_hline()

ggtitle()
```

Ensure histograms for each category that we are filling is overlaid side-by-side.

Axis Transformations

```
xlim(c(date(“”), date(“”)))
scale_x_continuous(break = c(..., ...))
scale_y_continuous()
scale_x_log10()
scale_y_log10()
```

Describing Visualization

Direction - Positive/Negative/little or no relationship
Strength
Shapen - Linear/nonlinear

Trends (lines) - Does a line describe a trend well?
Distributions (scatters, histogram) - How spread out is the data?
Distribtuion of two variables (scatters) - Clear / Strong or weak or no relationship?
Amounts (bars)

Important packages

ggplot2 - Creates visualizations
RColorBrewer - Pick color schemes
lubridate - Convert character strings to date vectors.

Styling the Plot

```
options(repr.plot.width = 9, repr.plot.height = 7)
ggplot(df, aes(x=..., y=..., colour=..., fill = as_factor(col)))

theme(text = element_text(size=12),
      axis.title = ...
      axis.text = ...
      title = ....
      legend.position= “top”,
      legend.direction = “vertical”)

ggplot(..., aes(x = ..., y = fct_reorder(column, criteria, .desc = TRUE))
```

Describing Visualization

```
ggsave("img/faithful_plot.png", faithful_plot)
ggsave("img/faithful_plot.jpg", faithful_plot)
ggsave("img/faithful_plot.bmp", faithful_plot)
ggsave("img/faithful_plot.tiff", faithful_plot)
ggsave("img/faithful_plot.svg", faithful_plot)
```

Facet Grid

```
facet_grid(rows = vars(column))
facet_grid(cols = vars(column))
```

Fill and Color

Fill

Used in geom_bar() and geom_histogram().
Aesthetically fills the bars.

Color

Used in geom_point() and geom_line()