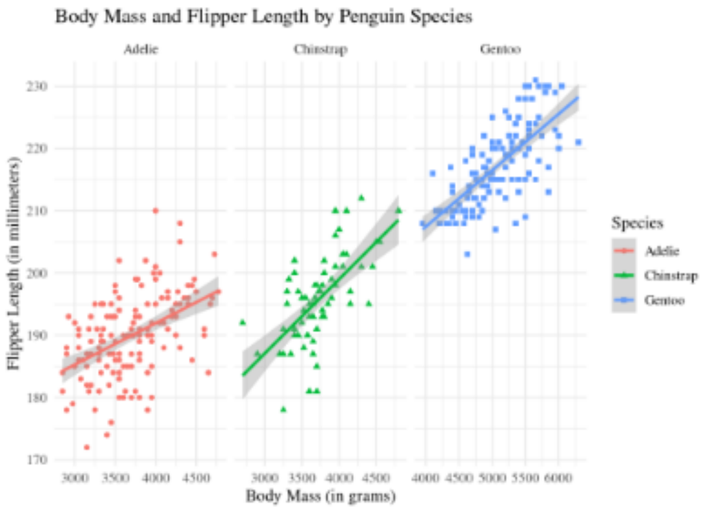
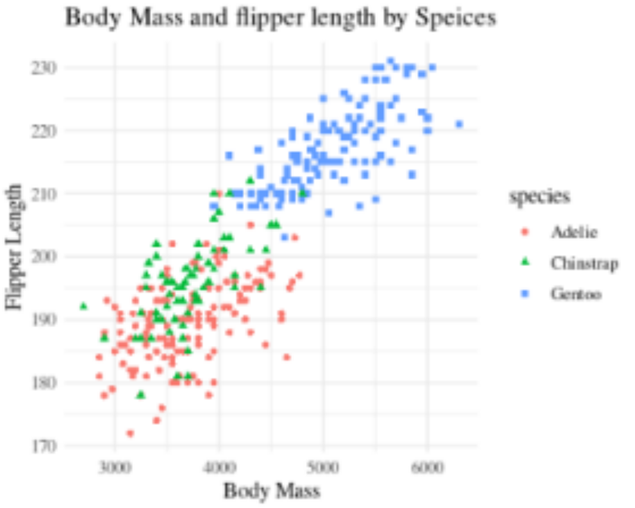
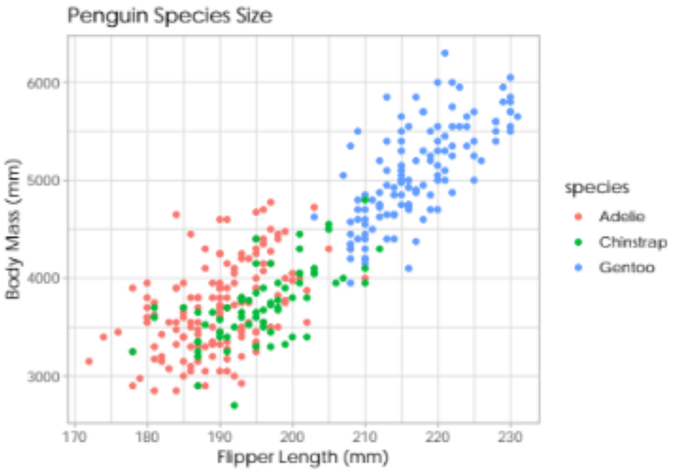
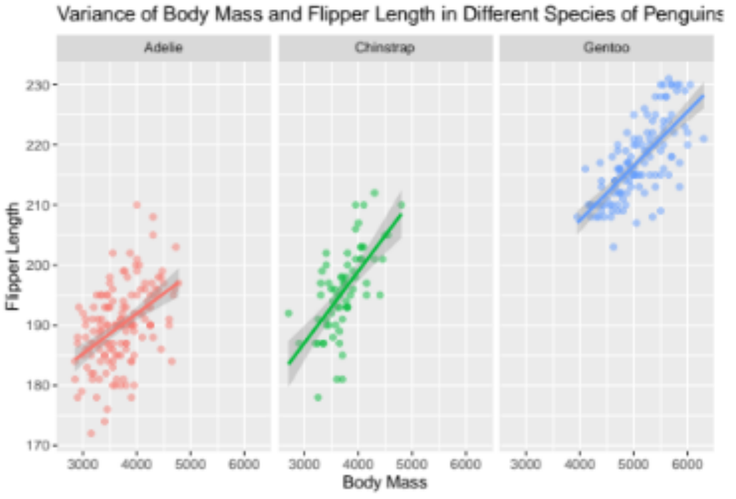
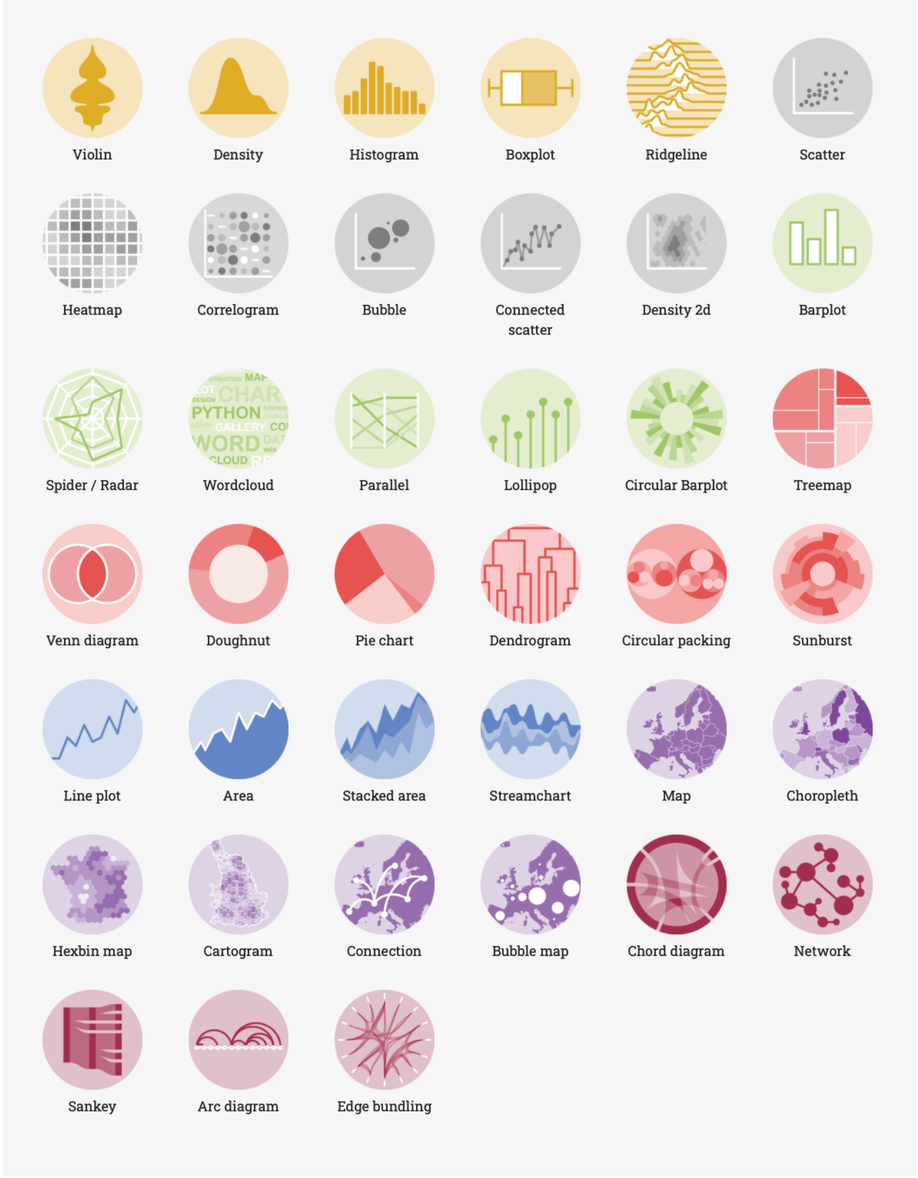

Body Mass and Flipper Length by Penguin Species
Adelie
Chinstrap
Gentoo
Flipper Length (in millimeters)
Body Mass (in grams)
Species
Adelie
Chinstrap
Gentoo


Body Mass and flipper length by Speices
Flipper Length
Body Mass
species
Adelie
Chinstrap
Gentoo

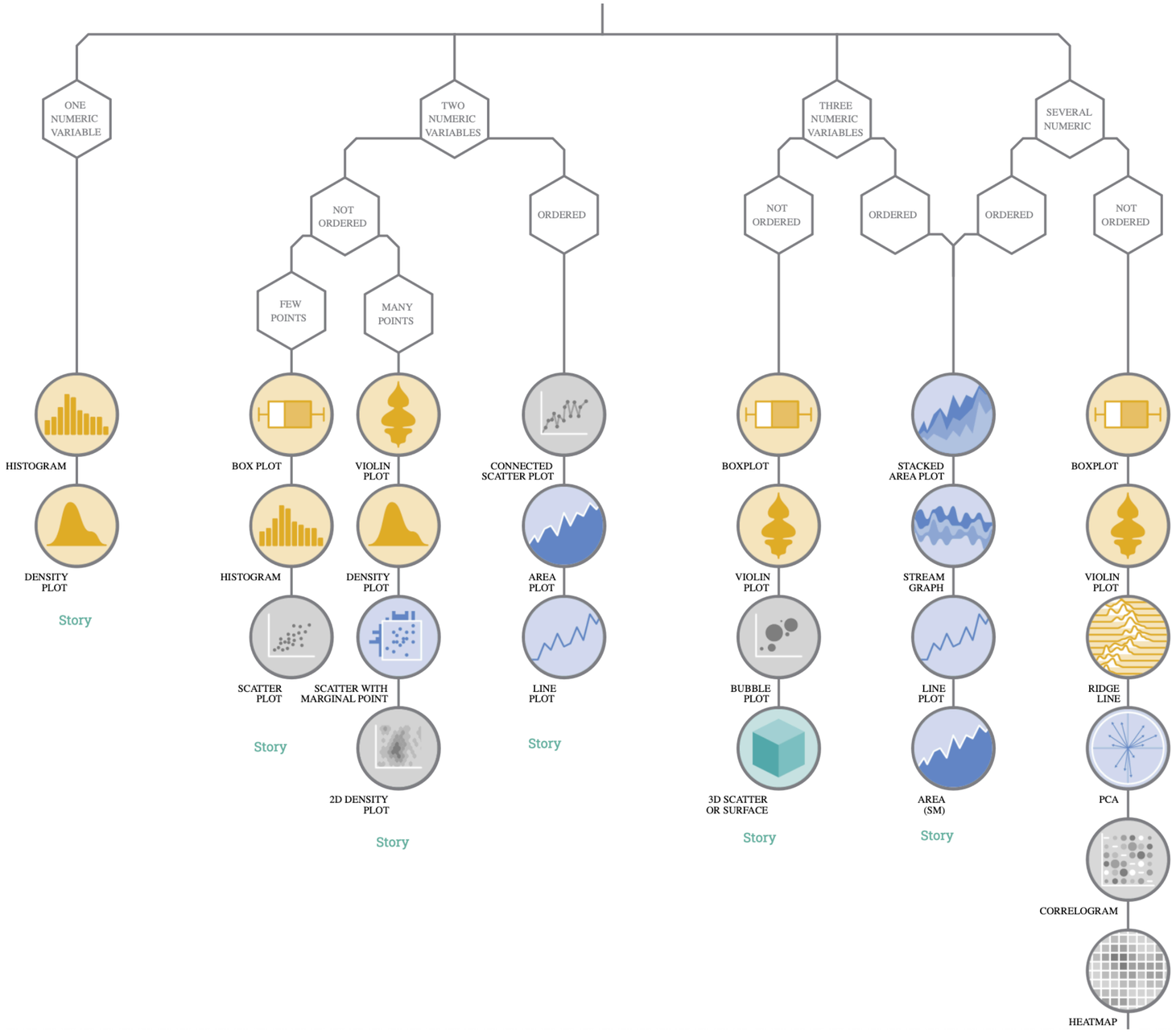

Penguin Species Size
Body Mass (mm)
Flipper Length (mm)
species
Adelie
Chinstrap
Gentoo


Variance of Body Mass and Flipper Length in Different Species of Penguins
Adelie
Chinstrap
Gentoo
Flipper Length
Body Mass

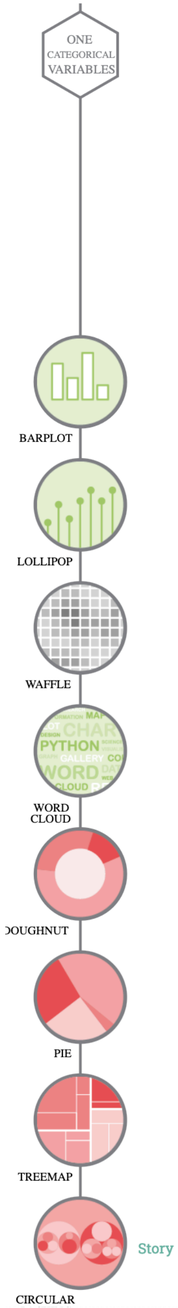
Types of Graphs

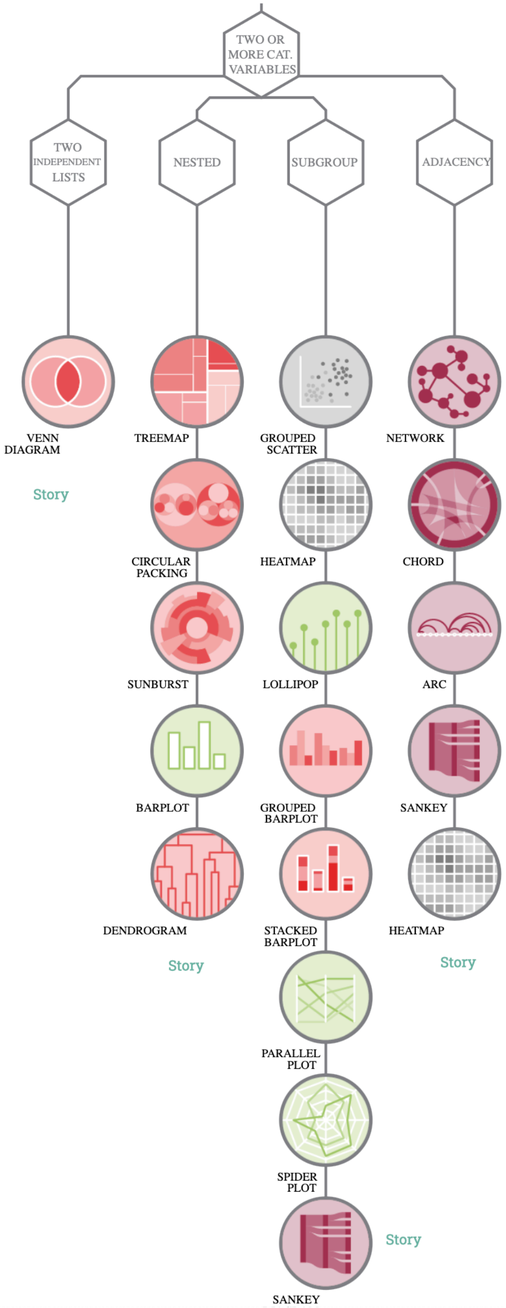


Numeric

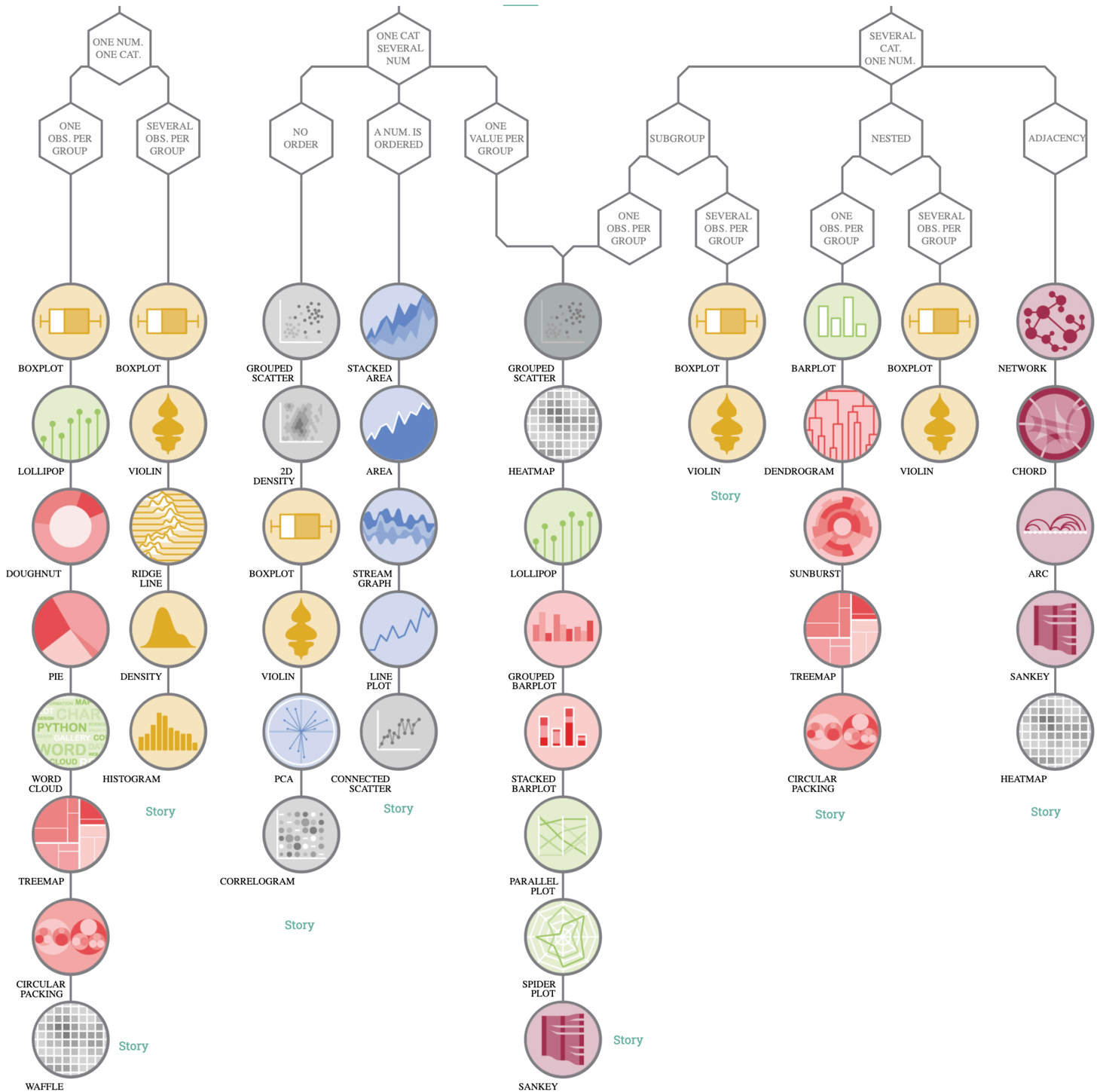

ONE NUMERIC VARIABLE
TWO NUMERIC VARIABLES
THREE NUMERIC VARIABLES
SEVERAL NUMERIC
NOT ORDERED
ORDERED
FEW POINTS
MANY POINTS
HISTOGRAM
DENSITY PLOT
Story
BOX PLOT
VIOLIN PLOT
HISTOGRAM
DENSITY PLOT
SCATTER PLOT
SCATTER WITH MARGINAL POINT
2D DENSITY PLOT
Story
CONNECTED SCATTER PLOT
AREA PLOT
LINE PLOT
Story
BOXPLOT
VIOLIN PLOT
BUBBLE PLOT
3D SCATTER OR SURFACE
Story
STACKED AREA PLOT
STREAM GRAPH
LINE PLOT
AREA (SM)
Story
BOXPLOT
VIOLIN PLOT
RIDGE LINE
PCA
CORRELOGRAM
HEATMAP

Categorical

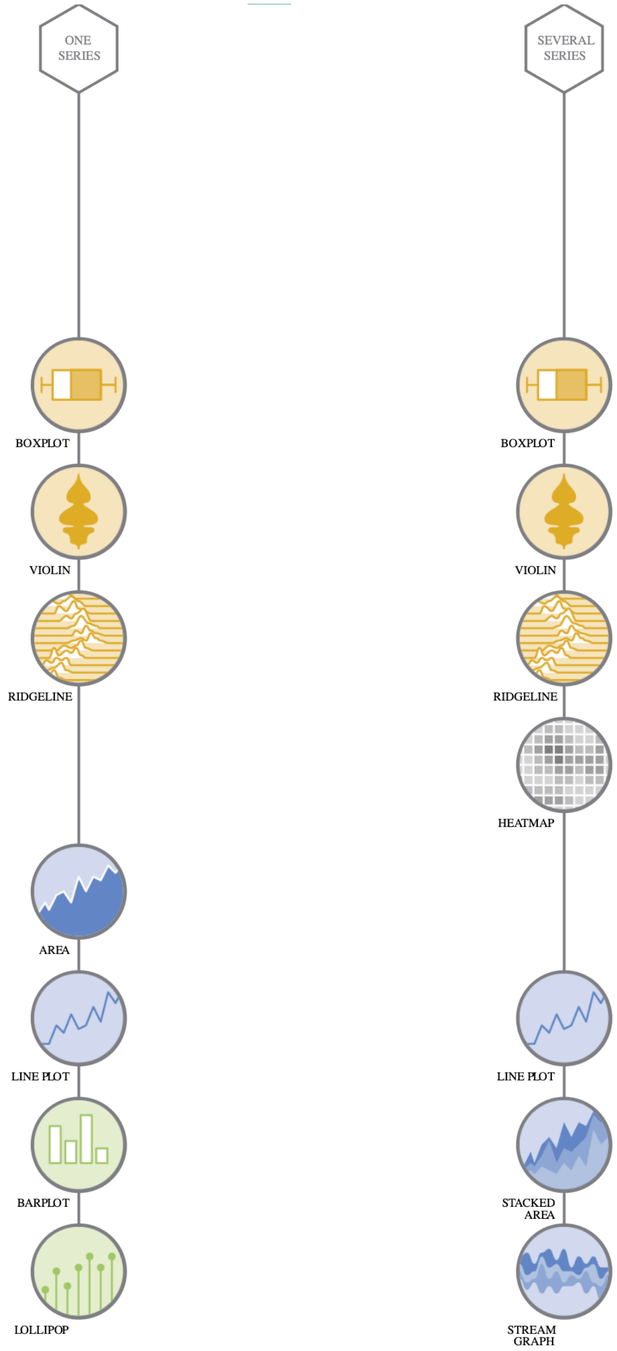




Numeric & Categorical



Time Series


ONE SERIES
BOXPLOT
VIOLIN
RIDGELINE
AREA
LINE PLOT
BARPLOT
LOLLIPOP
SEVERAL SERIES
BOXPLOT
VIOLIN
RIDGELINE
HEATMAP
LINE PLOT
STACKED AREA
STREAM GRAPH

MAPS

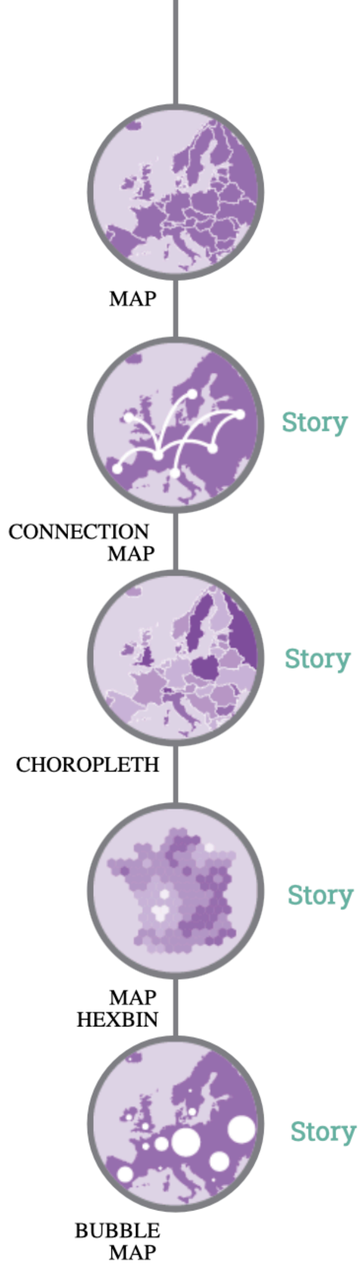
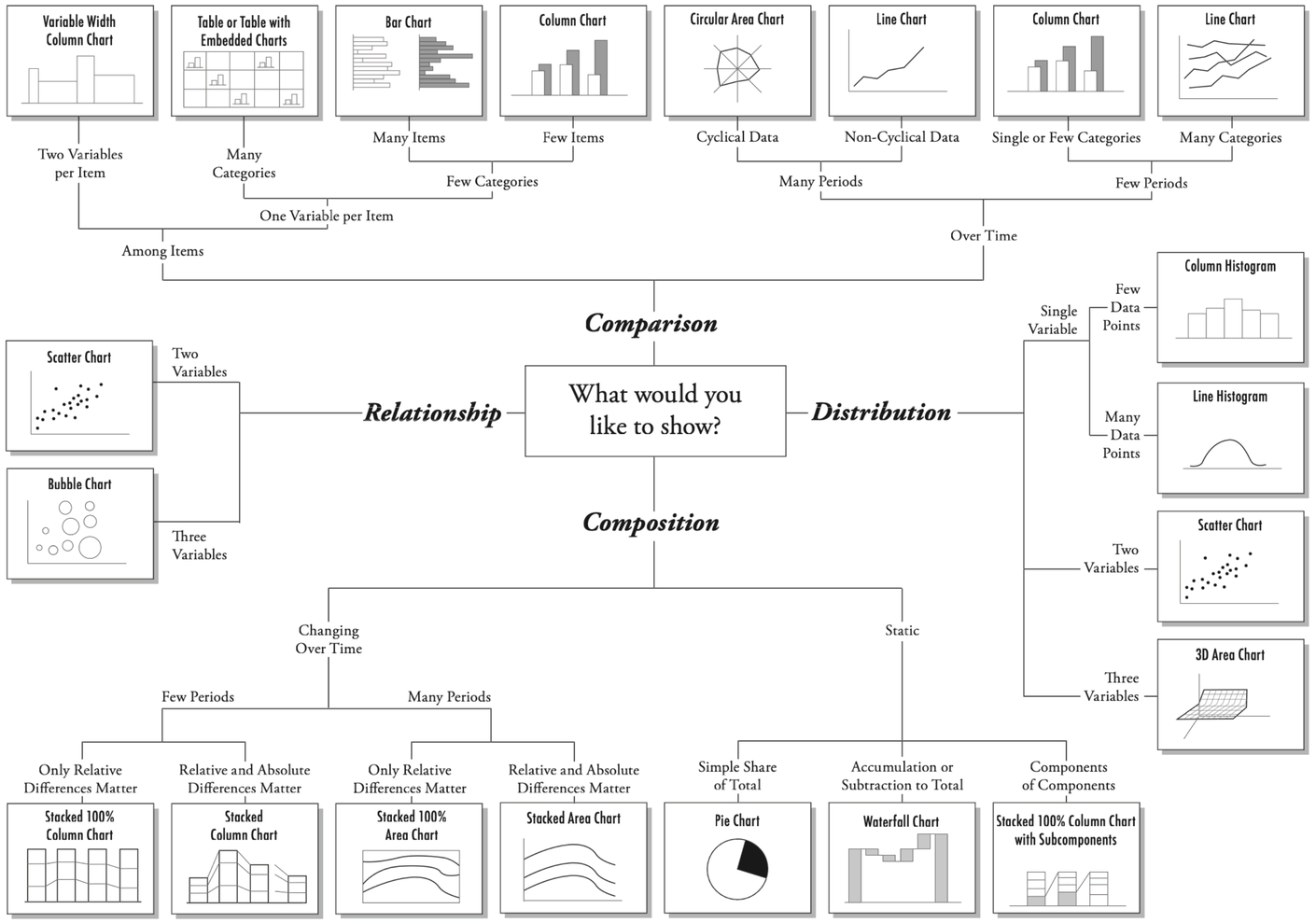
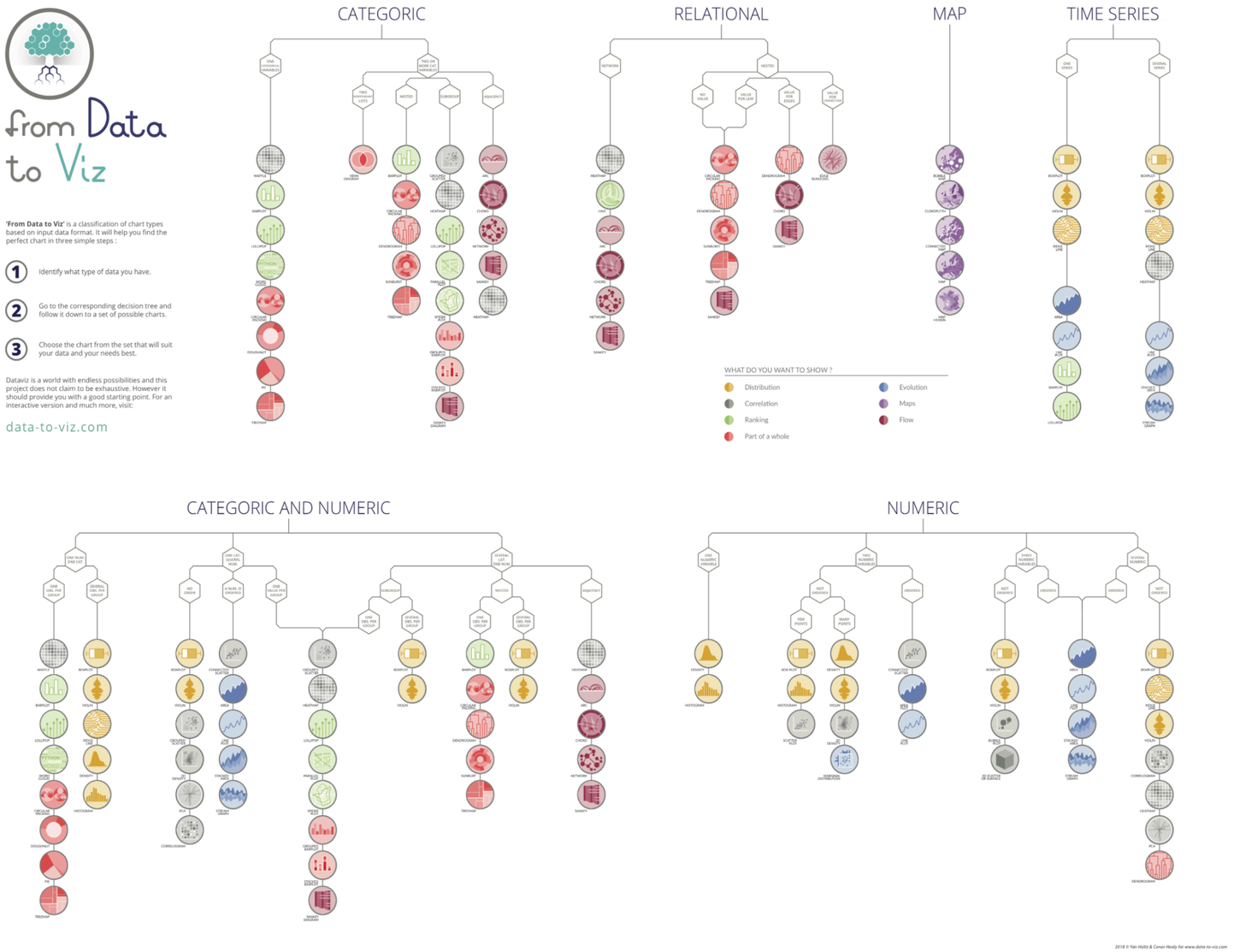

MAP
Story
CONNECTION MAP
Story
CHOROPLETH
Story
MAP HEXBIN
Story
BUBBLE MAP

Chart Suggestions—A Thought-Starter

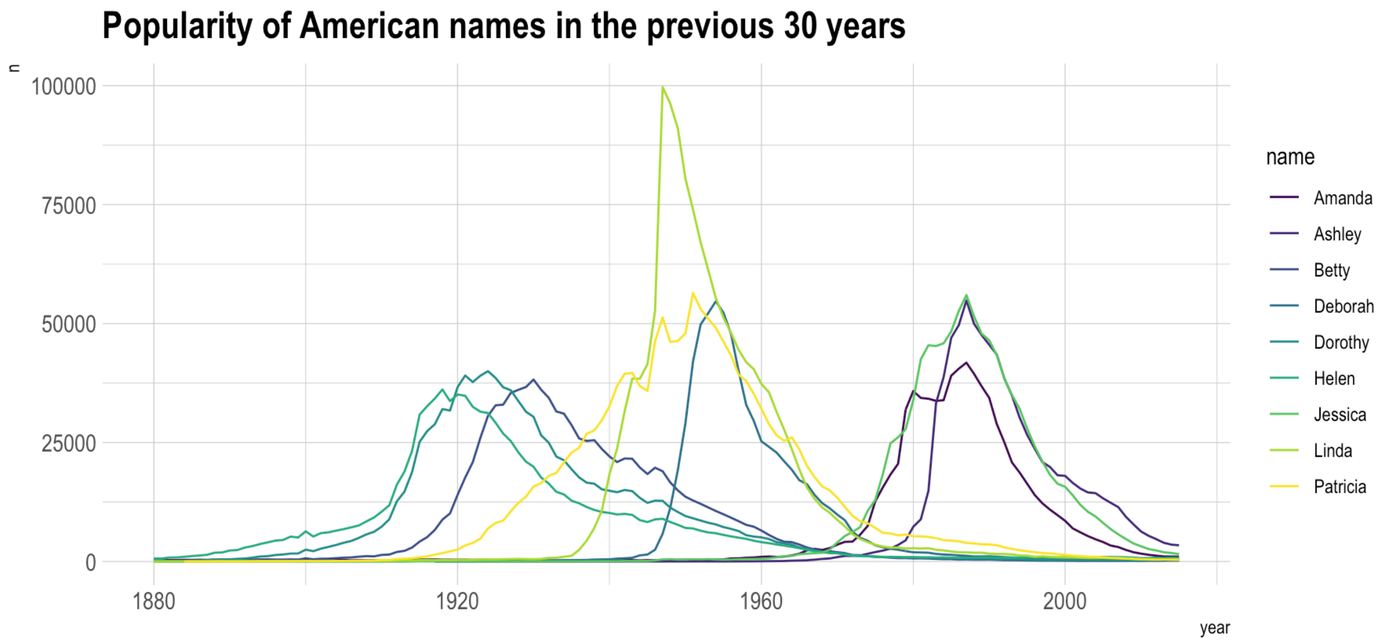




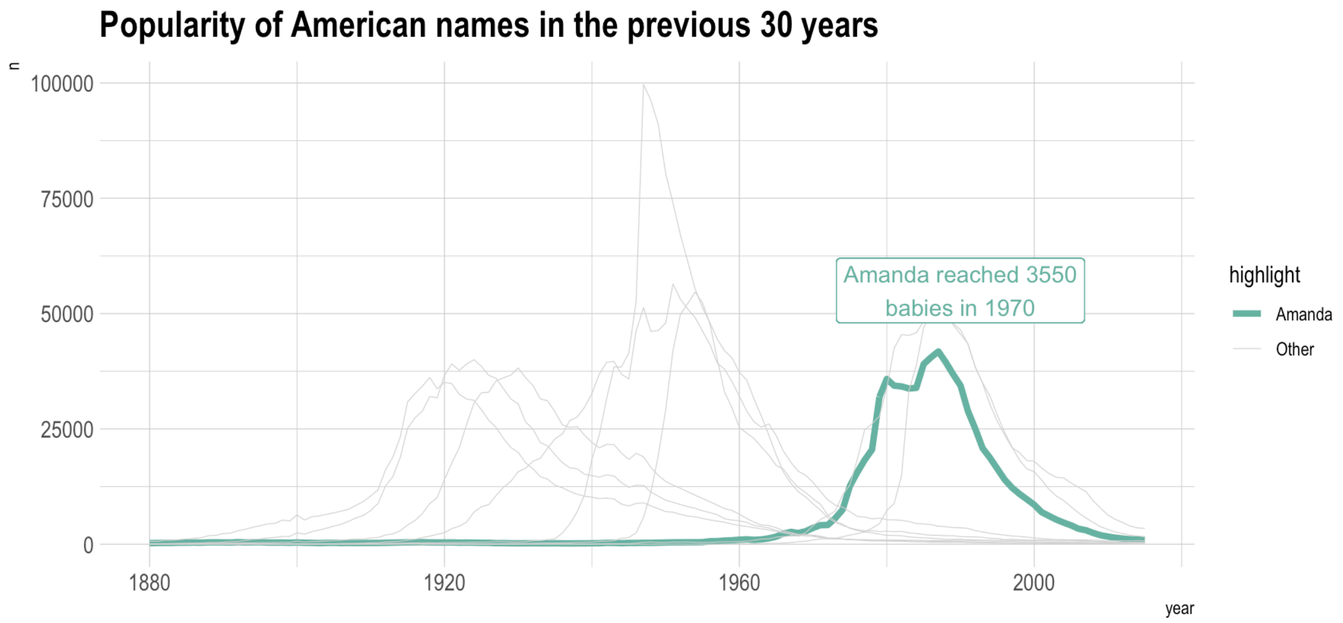
https://www.data-to-viz.com/

year	name	n
1880	Amanda	241
1880	Betty	117
1880	Dorothy	112
1880	Helen	636
1880	Linda	27

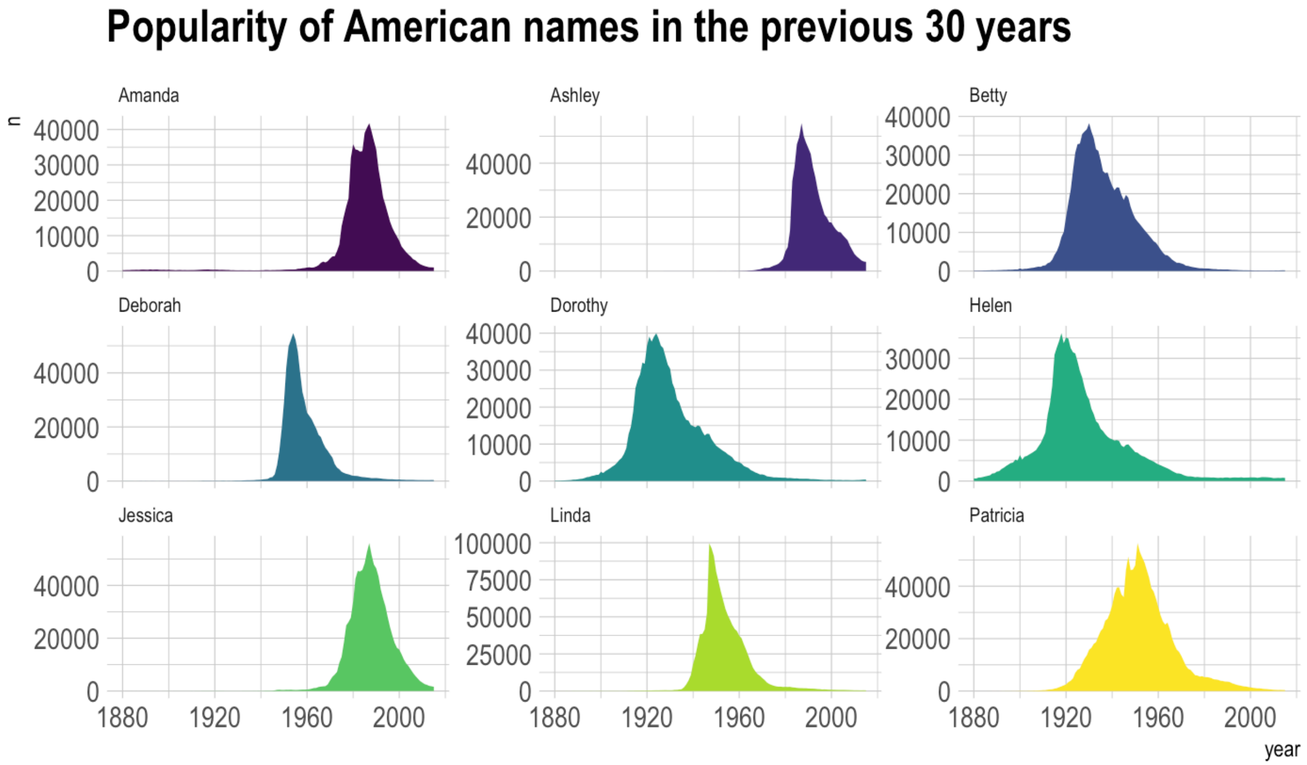
```r
data %>%
  ggplot( aes(x=year, y=n, group=name, color=name)) +
    geom_line() +
    scale_color_viridis(discrete = TRUE) +
    theme(legend.position="none") +
    ggtitle("Popularity of American names in the previous 30 years") +
    theme_ipsum()
```



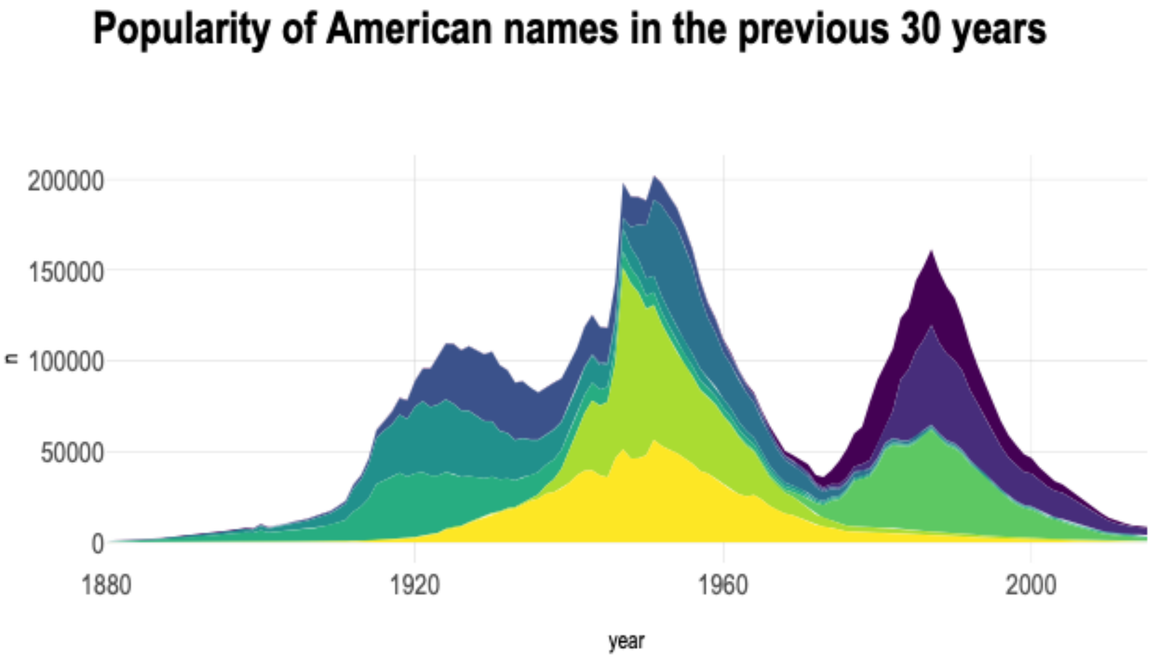
```r
data %>%
  mutate( highlight=ifelse(name=="Amanda", "Amanda", "Other")) %>%
  ggplot( aes(x=year, y=n, group=name, color=highlight, size=highlight)) +
    geom_line() +
    scale_color_manual(values = c("#69b3a2", "lightgrey")) +
    scale_size_manual(values=c(1.5,0.2)) +
    theme(legend.position="none") +
    ggtitle("Popularity of American names in the previous 30 years") +
    theme_ipsum() +
    geom_label( x=1990, y=55000, label="Amanda reached 3550\nbabies in 1970", size=4, color="#69b3a2")
```



```
data %>%
  ggplot( aes(x=year, y=n, group=name, fill=name)) +
    geom_area() +
    scale_fill_viridis(discrete = TRUE) +
    theme(legend.position="none") +
    ggtitle("Popularity of American names in the previous 30 years") +
    theme_ipsum() +
    theme(
      legend.position="none",
      panel.spacing = unit(0.1, "lines"),
      strip.text.x = element_text(size = 8)
    ) +
    facet_wrap(~name, scale="free_y")
```



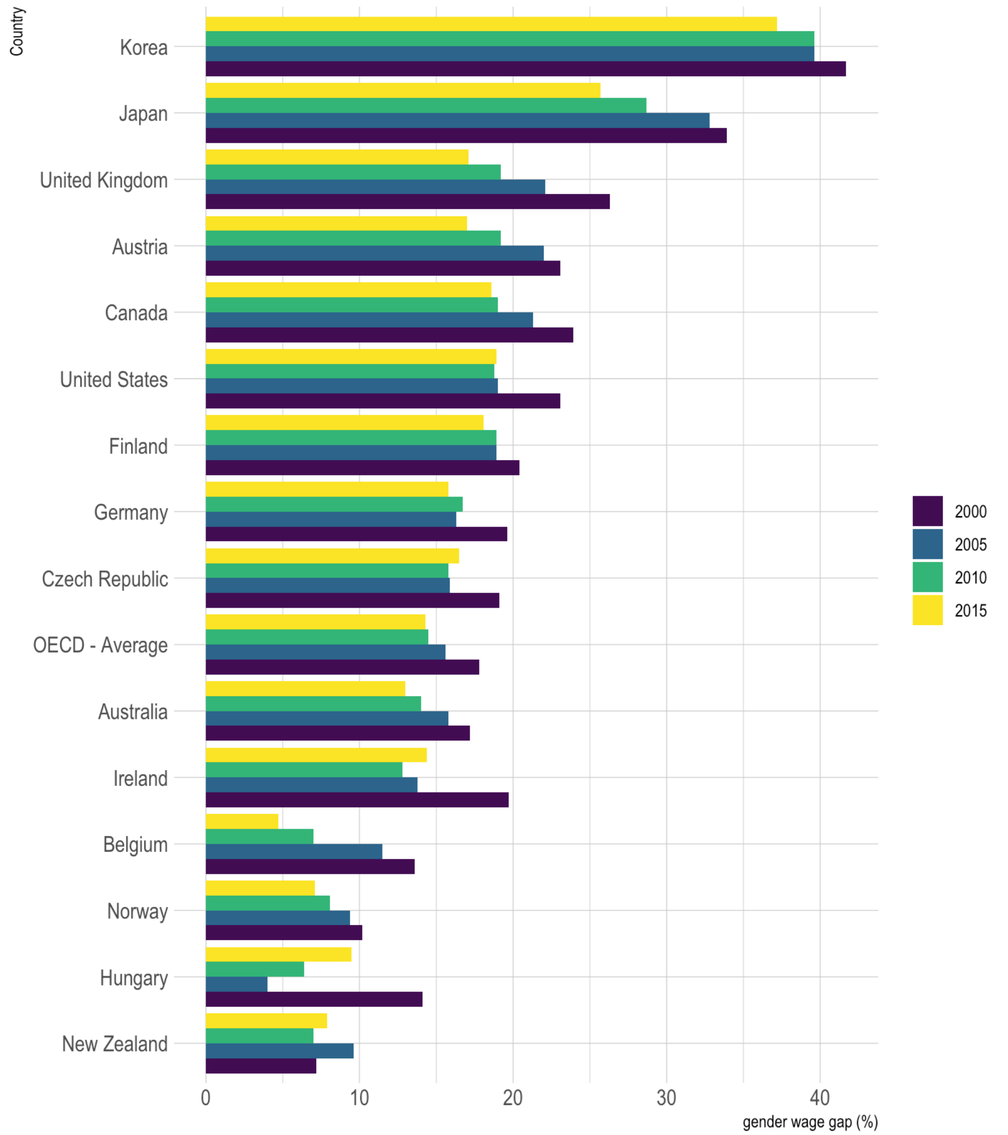
```r
p <- data %>%
  ggplot( aes(x=year, y=n, fill=name, text=name)) +
    geom_area( ) +
    scale_fill_viridis(discrete = TRUE) +
    theme(legend.position="none") +
    ggtitle("Popularity of American names in the previous 30 years") +
    theme_ipsum() +
    theme(legend.position="none")
ggplotly(p, tooltip="text")
```



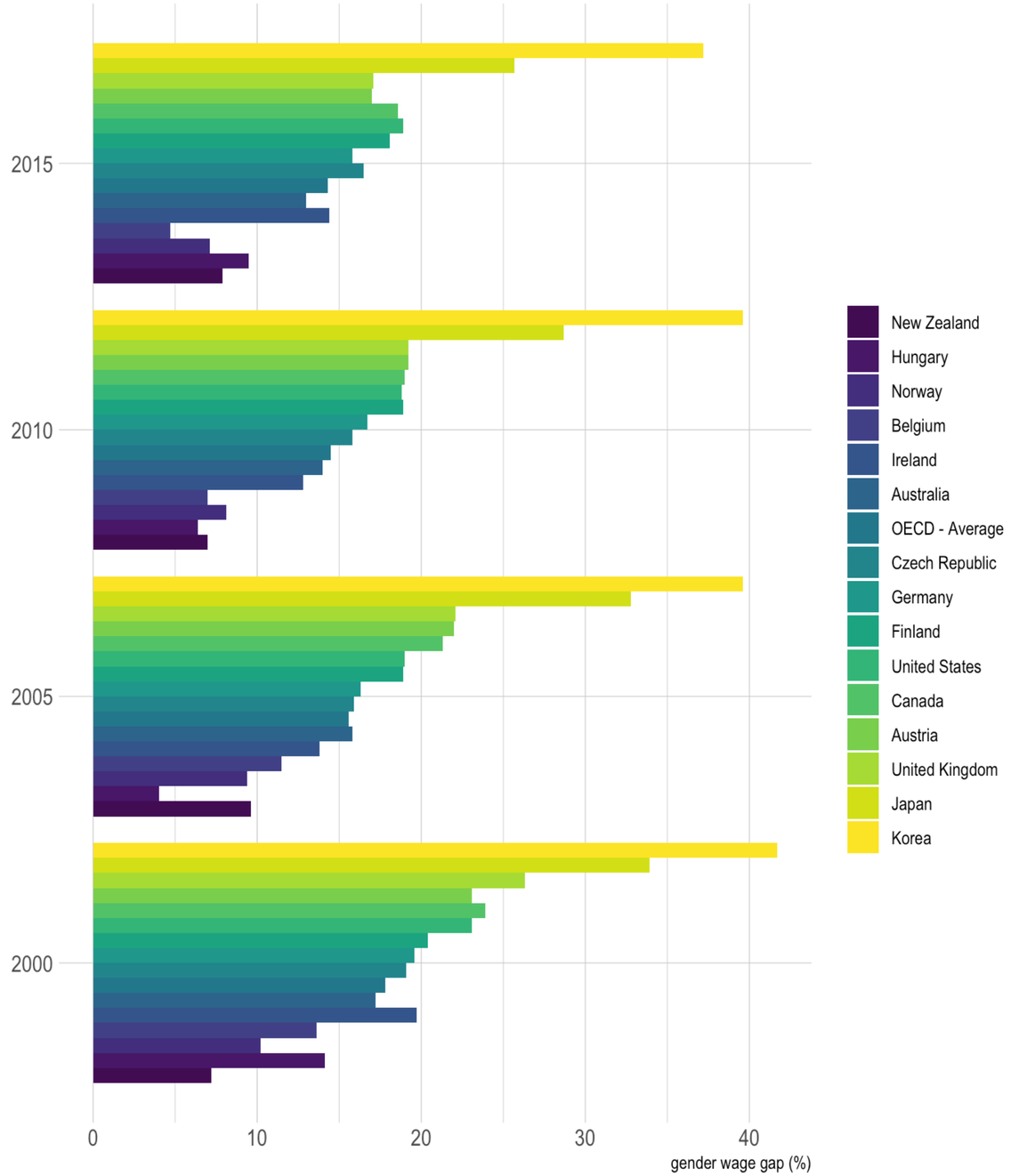
https://www.data-to-viz.com/story/OneCatSevOrderedNum.html

Country	TIME	Value
Australia	2000	17.2
Australia	2005	15.8
Australia	2010	14.0
Australia	2015	13.0
Austria	2000	23.1
Austria	2005	22.0

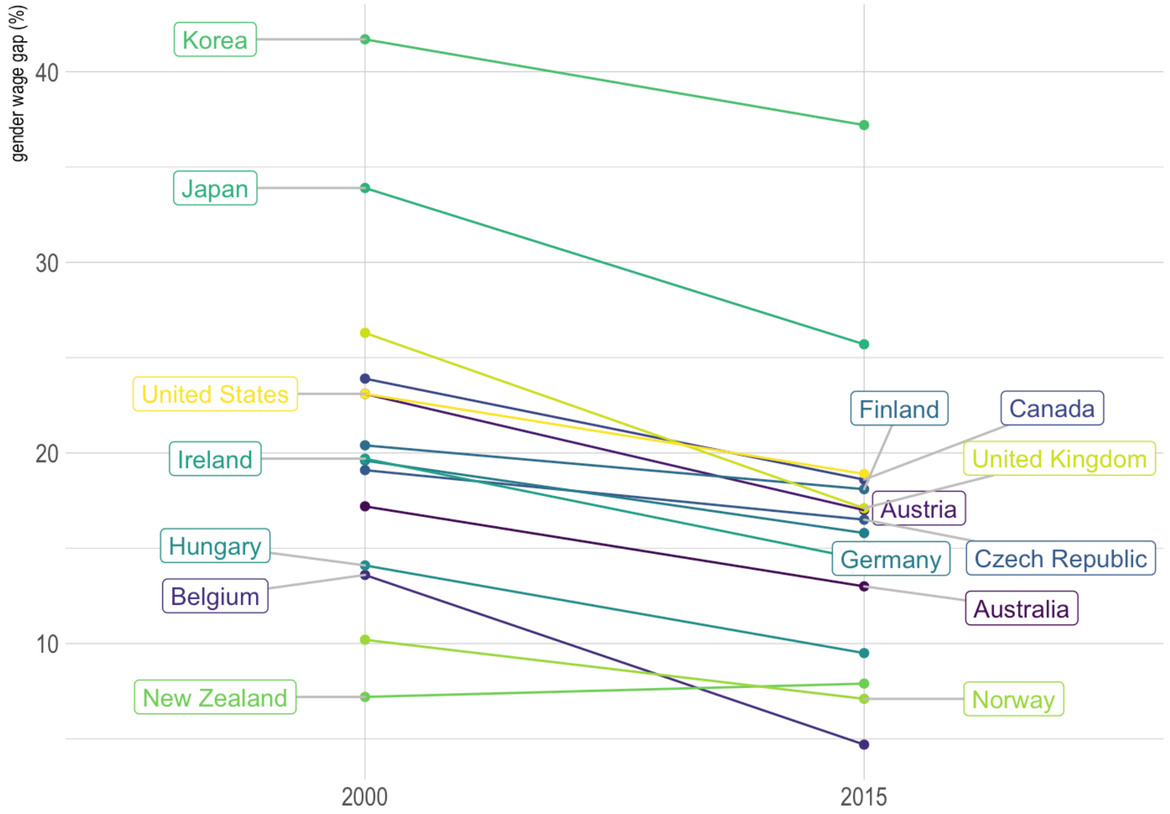
```r
# List of country with 4 values
with4 <- data %>%
  group_by(Country) %>%
  summarize(n=n()) %>%
  filter(n==4)

# Grouped
data %>%
  filter(Country %in% with4$Country) %>%
  mutate(Country = fct_reorder(Country, Value)) %>%
  mutate(TIME=factor(TIME, levels = c("2000", "2005", "2010", "2015"))) %>%
  ggplot(aes(fill=as.factor(TIME), y=Value, x=Country)) +
    geom_bar(position="dodge", stat="identity") +
    scale_fill_viridis(discrete=T, name="") +
    coord_flip() +
    theme_ipsum()  +
    ylab("gender wage gap (%)")
```

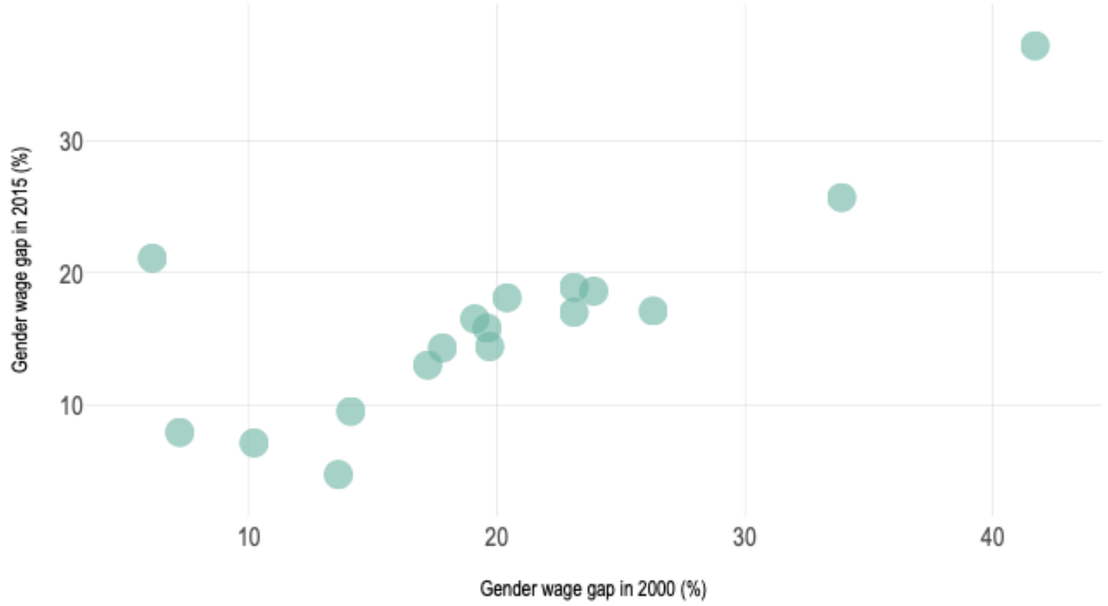


```r
# Grouped
data %>%
  filter(Country %in% with4$Country) %>%
  mutate(Country = fct_reorder(Country, Value)) %>%
  mutate(TIME=factor(TIME, levels = c("2000", "2005", "2010", "2015"))) %>%
  ggplot(aes(fill=Country, y=Value, x=as.factor(TIME))) +
    geom_bar(position="dodge", stat="identity") +
    scale_fill_viridis(discrete=T, name="") +
    coord_flip() +
    theme_ipsum() +
    xlab("") +
    ylab("gender wage gap (%)")
```



```r
# Groups
all <- unique(data$Country)
grp1 <- sample( all, 20)
grp2 <- all[ ! all%in%grp1]

# Grouped
data %>%
  filter(Country %in% with4$Country) %>%
  filter(Country != "OECD - Average") %>%
  filter(TIME %in% c(2000, 2015)) %>%
  mutate(label = if_else(TIME == max(TIME) & Country %in% grp1, as.character(Country), NA_character_)) %>%
  mutate(label2 = if_else(TIME == min(TIME) & Country %in% grp2, as.character(Country), NA_character_)) %>%
  ggplot( aes(x=as.factor(TIME), y=Value, color=Country, group=Country)) +
    geom_point() +
    geom_line() +
    geom_label_repel( aes(label=label), nudge_x = 0.3, hjust=0, na.rm = TRUE, segment.colour="grey") +
    geom_label_repel( aes(label=label2), nudge_x = -0.3, hjust=1, na.rm = TRUE, segment.colour="grey") +
    scale_color_viridis(discrete=T, name="") +
    theme_ipsum() +
    theme(
      legend.position ="none"
    ) +
    xlab("") +
    ylab("gender wage gap (%)")
```



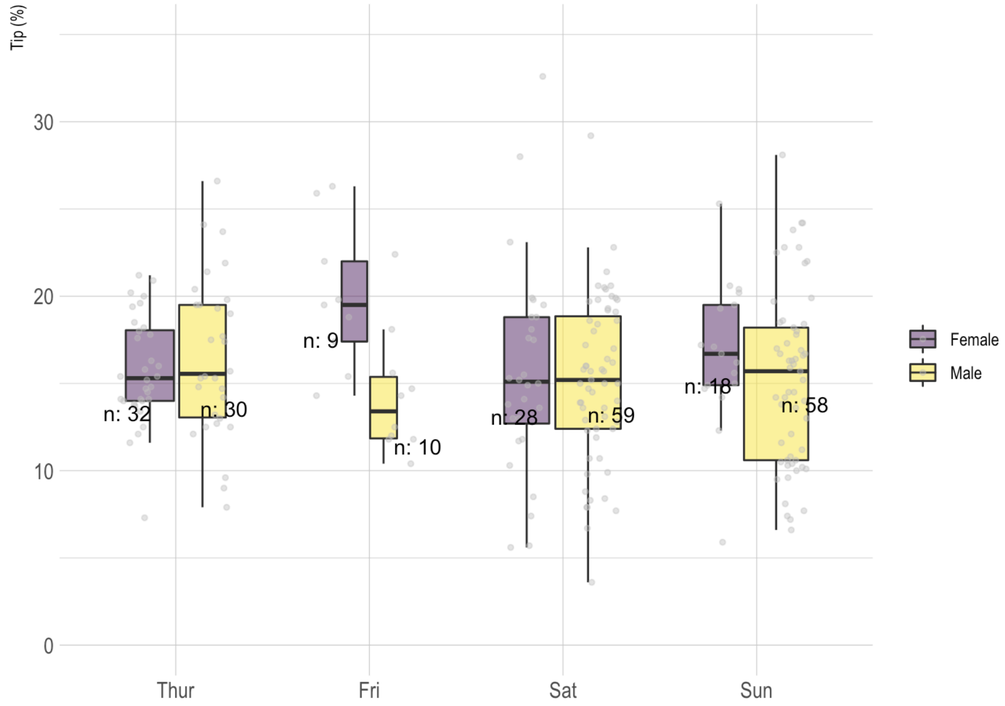



```
p <- data %>%
  filter(TIME %in% c(2000, 2015)) %>%
  spread(key=TIME, value=Value, -1) %>%
  filter(`2000`!=-1) %>%
  filter(`2015`!=-1) %>%
  mutate(text=paste("Country: ",Country, "\n", "Wage gap in 2000: ", `2000`, "%\nWage gap in 2015: ", `2015`,
sep="")) %>%
  ggplot(aes(x=`2000`, y=`2015`, text=text)) +
    geom_point(size=4, alpha=0.6, color="#69b3a2") +
    theme_ipsum() +
    theme(legend.position="none") +
    xlab("Gender wage gap in 2000 (%)") +
    ylab("Gender wage gap in 2015 (%)")


ggplotly(p, tooltip="text")
```

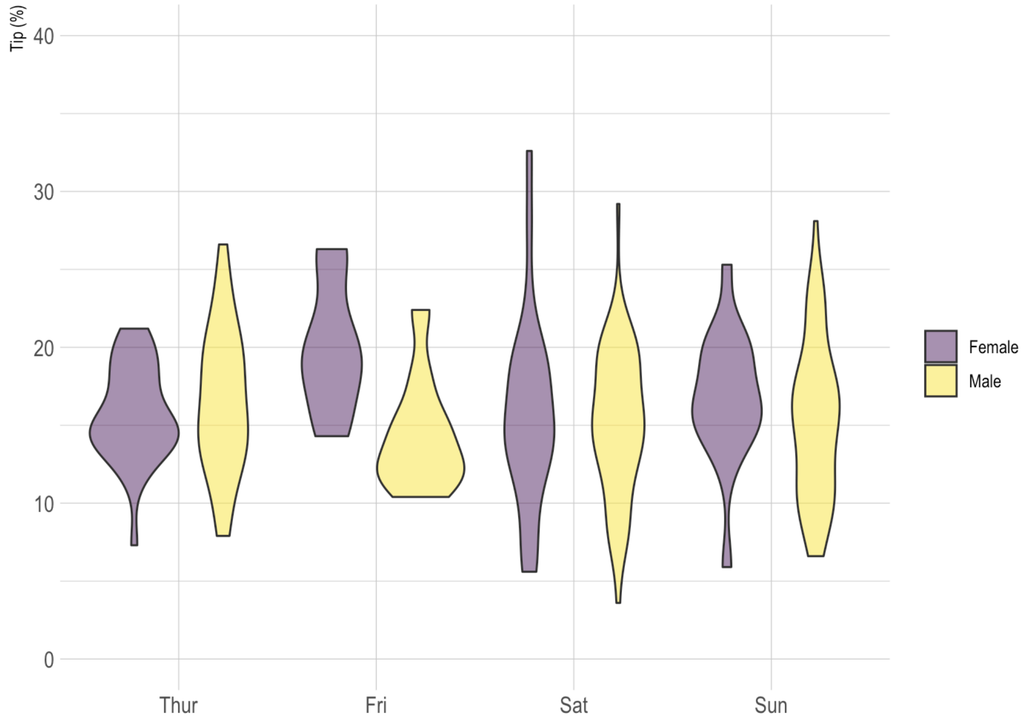
https://www.data-to-viz.com/story/OneNumSevCatSubgroupOneObsPerGroup.html

tip	sex	day	smoker
5.9	Female	Sun	No
16.1	Male	Sun	No
16.7	Male	Sun	No
14.0	Male	Sun	No
14.7	Female	Sun	No
18.6	Male	Sun	No

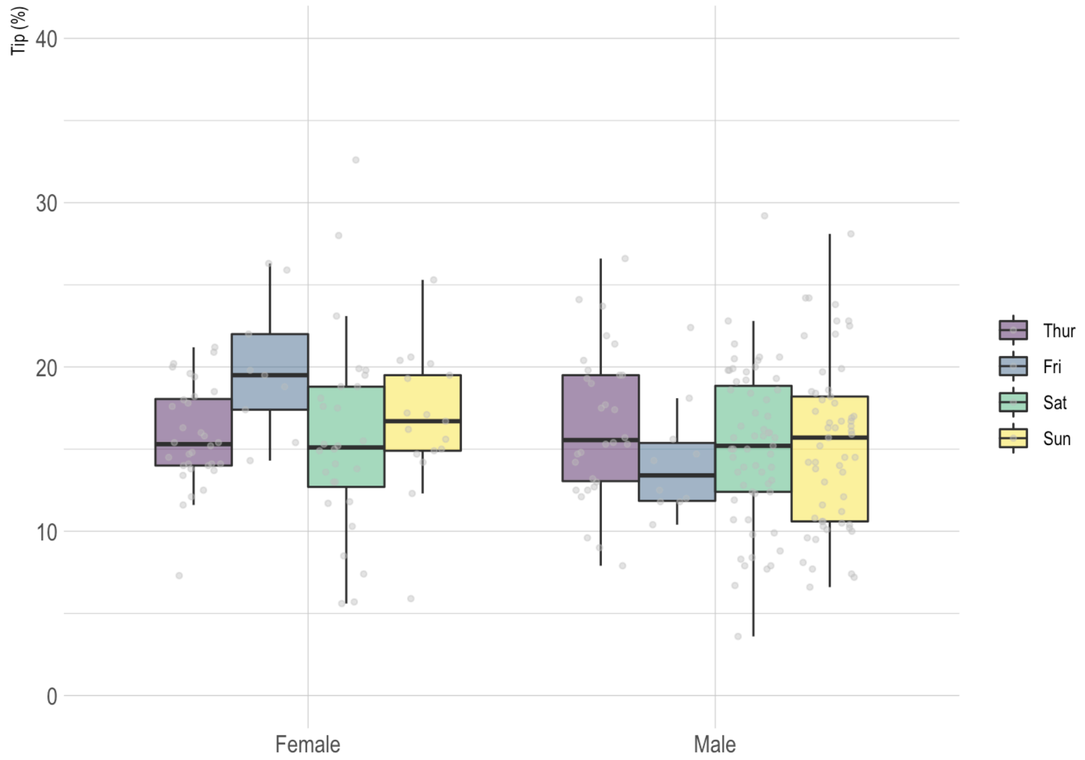
```r
# Counts the number of value per group and subgroup
counts = data %>%
  group_by(day, sex) %>%
  summarize(
    n=n(),
    median=median(tip)
  )

# Grouped
data %>%
  mutate(day = fct_reorder(day, tip)) %>%
  mutate(day = factor(day, levels=c("Thur", "Fri", "Sat", "Sun"))) %>%
  ggplot(aes(fill=sex, y=tip, x=day)) +
    geom_boxplot(position=position_dodge2(preserve = "total"), alpha=0.5, outlier.colour="transparent", varwi
dth = TRUE) +
    geom_point(color="grey", size=1, width=0.1, position=position_jitterdodge() , alpha=0.4) +
    scale_fill_viridis(discrete=T, name="") +
    geom_text(data=counts, aes(label=paste0("n: ",n), y=median-2), position=position_dodge(1), hjust=0.5) +
    theme_ipsum()  +
    xlab("") +
    ylab("Tip (%)") +
    ylim(0,35)
```



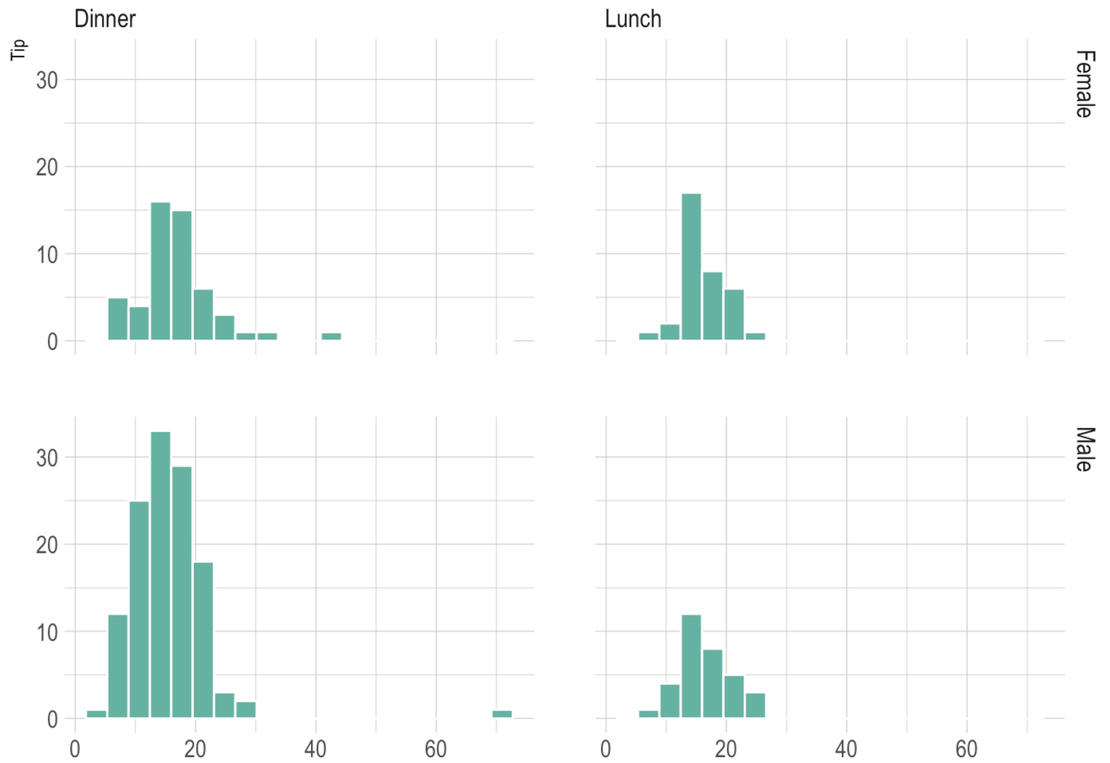
```r
# Grouped
data %>%
  mutate(day = fct_reorder(day, tip)) %>%
  mutate(day = factor(day, levels=c("Thur", "Fri", "Sat", "Sun"))) %>%
  ggplot(aes(fill=sex, y=tip, x=day)) +
    geom_violin(position="dodge", alpha=0.5, outlier.colour="transparent") +
    scale_fill_viridis(discrete=T, name="") +
    theme_ipsum()  +
    xlab("") +
    ylab("Tip (%)") +
    ylim(0,40)
```



```r
# Grouped
data %>%
  mutate(day = fct_reorder(day, tip)) %>%
  mutate(day = factor(day, levels=c("Thur", "Fri", "Sat", "Sun"))) %>%
  ggplot(aes(fill=day, y=tip, x=sex)) +
    geom_boxplot(position="dodge", alpha=0.5, outlier.colour="transparent") +
    geom_point(color="grey", size=1, width=0.1, position=position_jitterdodge() , alpha=0.4) +
    scale_fill_viridis(discrete=T, name="") +
    theme_ipsum()  +
    xlab("") +
    ylab("Tip (%)") +
    ylim(0,40)
```

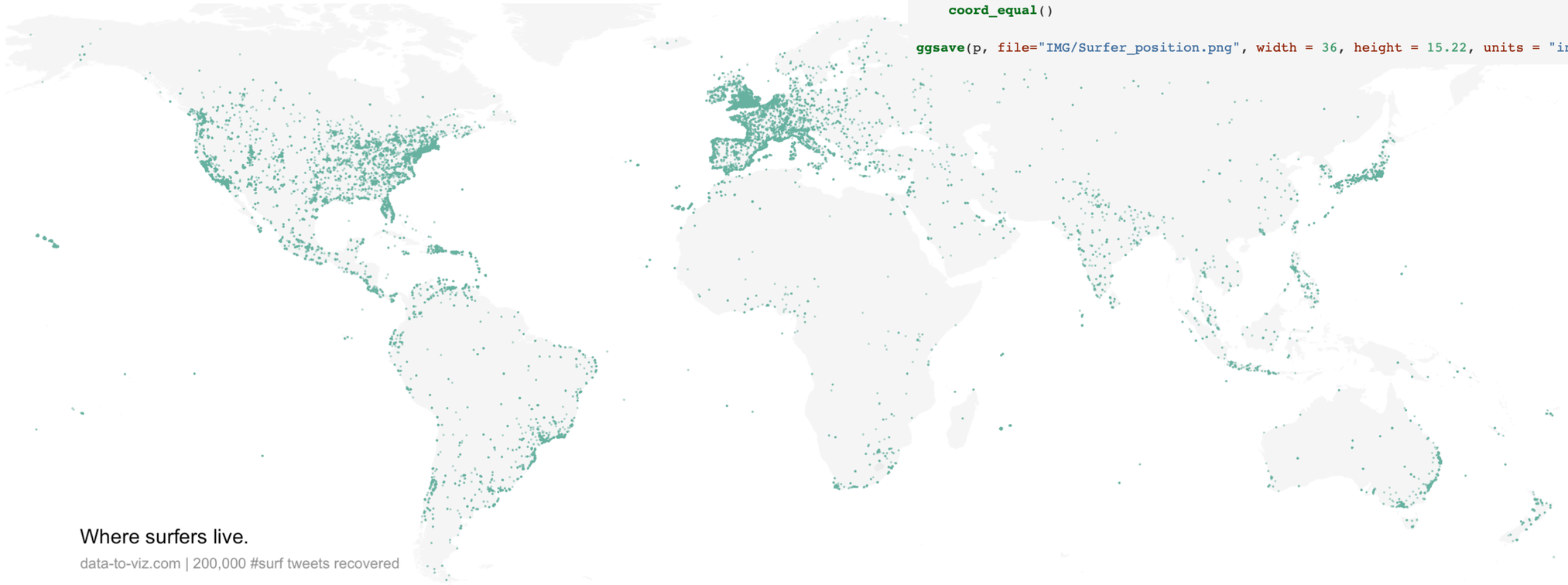


```r
# Grouped
data %>%
  mutate(day = fct_reorder(day, tip)) %>%
  mutate(day = factor(day, levels=c("Thur", "Fri", "Sat", "Sun"))) %>%
  ggplot(aes(x=tip)) +
    geom_histogram(bins=20, fill="#69b3a2", color="white") +
    facet_grid(sex~time) +
    theme_ipsum()  +
    xlab("") +
    ylab("Tip")
```



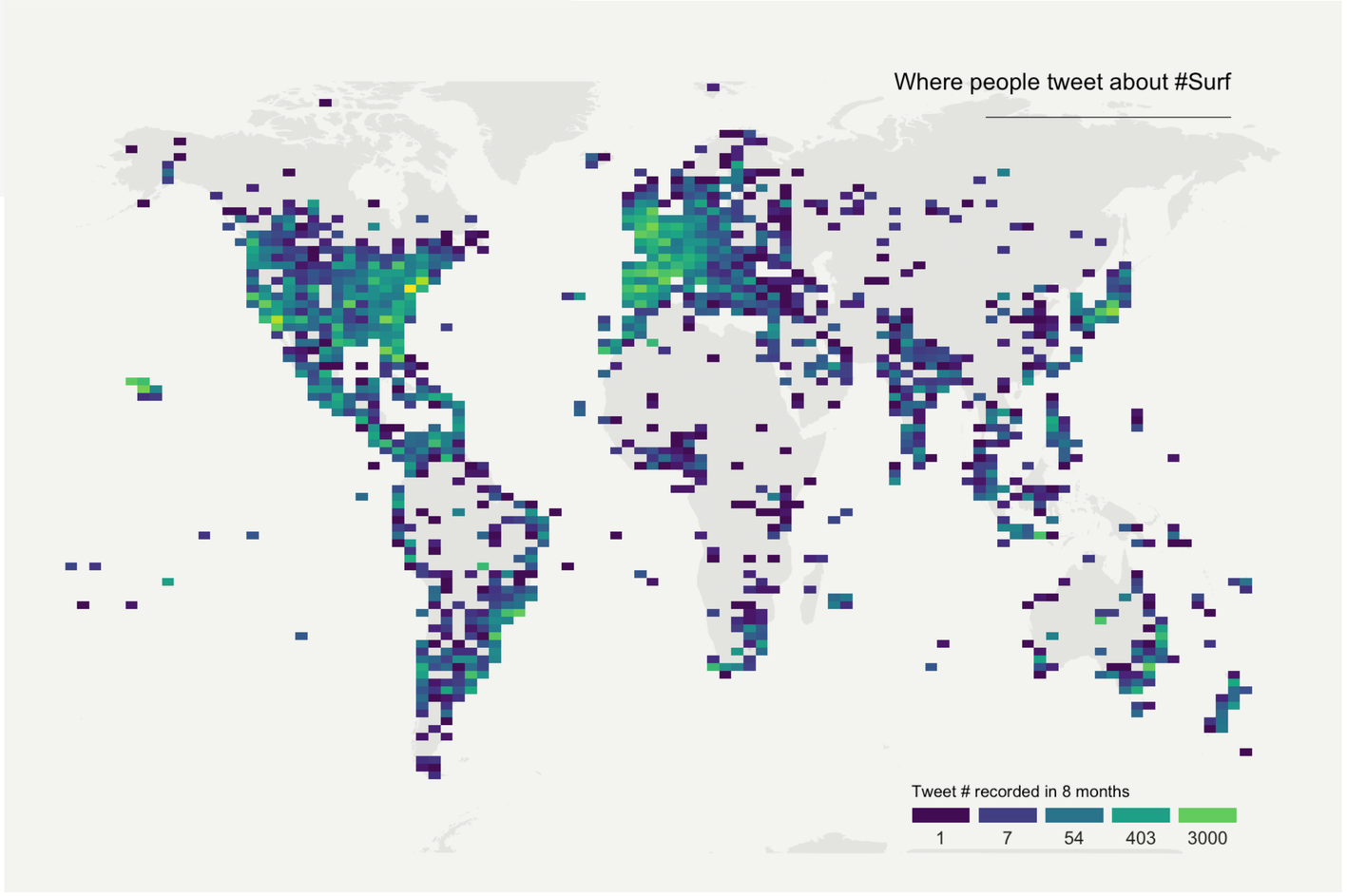
homelat	homelon	homecontinent	travellat	travellon	travelcontinent	n
-53.77363	-67.70729	South America	-38.008100	-57.539994	South America	1
-44.88564	168.66616	Australia	-44.500000	169.283333	Australia	1
-42.92278	147.41639	*NA*	-8.667070	115.139337	Asia	1
-42.76945	-65.03172	South America	-42.773494	-65.027336	South America	1
-42.73562	147.43785	Australia	-42.991864	147.524385	Australia	1
-42.40082	173.68139	*NA*	30.503338	-9.680982	Africa	1
-42.40082	173.68139	*NA*	30.514908	-9.680239	Africa	1
-42.40082	173.68139	*NA*	30.544799	-9.710519	Africa	1
-42.40082	173.68139	*NA*	30.545095	-9.709994	Africa	1
-42.40082	173.68139	*NA*	43.661345	-1.435023	Europe	1
-42.40082	173.68139	*NA*	54.270005	-8.609789	*NA*	1
-42.40082	173.68139	*NA*	54.270247	-8.609173	*NA*	1
-42.40082	173.68139	*NA*	54.271900	-8.593330	*NA*	1
-41.46892	-72.94114	South America	-4.109925	-81.064500	South America	1
-41.46892	-72.94114	South America	-4.109823	-81.063089	South America	1
-41.28646	174.77624	Australia	-41.333300	174.783000	*NA*	1
-41.28646	174.77624	Australia	-39.116667	177.200000	Australia	1

```r
# Get the world polygon
world <- map_data("world")

p <- data %>%
  #head(1000) %>%
  ggplot( aes(x=homelon, y=homelat)) +
    geom_polygon(data = world, aes(x=long, y = lat, group = group), fill="grey", alpha=0.1) +
    geom_point(size=0.8, color="#69b3a2", alpha=0.5) +
    coord_equal() +
    theme_void() +
    theme(
        panel.spacing=unit(c(0,0,0,0), "null"),
        plot.margin=grid::unit(c(0,0,0,0), "cm"),
    ) +
    ggplot2::annotate("text", x = -150, y = -45, hjust = 0, size = 11, label = paste("Where surfers live."),
color = "Black") +
    ggplot2::annotate("text", x = -150, y = -51, hjust = 0, size = 8, label = paste("data-to-viz.com | 200,00
0 #surf tweets recovered"), color = "black", alpha = 0.5) +
    xlim(-180,180) +
    ylim(-60,80) +
    scale_x_continuous(expand = c(0.006, 0.006)) +
    coord_equal()

ggsave(p, file="IMG/Surfer_position.png", width = 36, height = 15.22, units = "in", dpi = 90)
```



```r
data %>%
  filter(homecontinent=='Europe') %>%
  ggplot( aes(x=homelon, y=homelat)) +
    geom_hex(bins=59) +
    ggplot2::annotate("text", x = -27, y = 72, label="Where people tweet about #Surf", colour = "black", size=5, alpha=1, hjust=0) +
    ggplot2::annotate("segment", x = -27, xend = 10, y = 70, yend = 70, colour = "black", size=0.2, alpha=1) +
    theme_void() +
    xlim(-30, 70) +
    ylim(24, 72) +
    scale_fill_viridis(
      trans = "log",
      breaks = c(1,7,54,403,3000),
      name="Tweet # recorded in 8 months",
      guide = guide_legend( keyheight = unit(2.5, units = "mm"), keywidth=unit(10, units = "mm"), label.position = "bottom", title.position = 'top', nrow=1)
    )  +
    ggtitle( "" ) +
    theme(
      legend.position = c(0.8, 0.09),
      legend.title=element_text(color="black", size=8),
      text = element_text(color = "#22211d"),
      plot.background = element_rect(fill = "#f5f5f2", color = NA),
      panel.background = element_rect(fill = "#f5f5f2", color = NA),
      legend.background = element_rect(fill = "#f5f5f2", color = NA),
      plot.title = element_text(size= 13, hjust=0.1, color = "#4e4d47", margin = margin(b = -0.1, t = 0.4, l = 2, unit = "cm")),
    )
```

```r
# Make the hexbin map with the geom_hex function
ggplot(data, aes(x=homelon, y=homelat)) +
    geom_polygon(data = world, aes(x=long, y = lat, group = group), fill="grey", alpha=0.3) +
    geom_bin2d(bins=100) +
    ggplot2::annotate("text", x = 175, y = 80, label="Where people tweet about #Surf", colour = "black", size=4, alpha=1, hjust=1) +
    ggplot2::annotate("segment", x = 100, xend = 175, y = 73, yend = 73, colour = "black", size=0.2, alpha=1) +
    theme_void() +
    ylim(-70, 80) +
    scale_fill_viridis(
      trans = "log",
      breaks = c(1,7,54,403,3000),
      name="Tweet # recorded in 8 months",
      guide = guide_legend( keyheight = unit(2.5, units = "mm"), keywidth=unit(10, units = "mm"), label.position = "bottom", title.position = 'top', nrow=1)
    )  +
    ggtitle( "" ) +
    theme(
      legend.position = c(0.8, 0.09),
      legend.title=element_text(color="black", size=8),
      text = element_text(color = "#22211d"),
      plot.background = element_rect(fill = "#f5f5f2", color = NA),
      panel.background = element_rect(fill = "#f5f5f2", color = NA),
      legend.background = element_rect(fill = "#f5f5f2", color = NA),
      plot.title = element_text(size= 13, hjust=0.1, color = "#4e4d47", margin = margin(b = -0.1, t =
    )
```



SWD BOOK
CHAPTER 1

The IMPORTANCE of CONTEXT

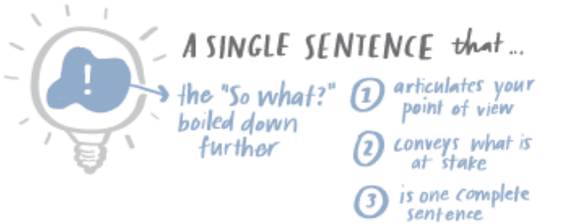
TYPES of ANALYSIS

WHERE to BEGIN?

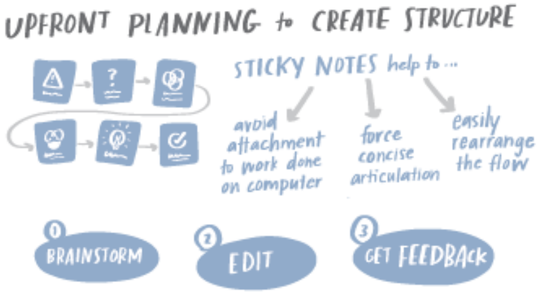
THREE MINUTE STORY

BIG IDEA*

*from Nancy Duarte (Resonate)



STORY BOARDing



SWD BOOK

The STORYTELLING with DATA PROCESS

STEP 1: UNDERSTAND the CONTEXT

- WHO is your audience?
- WHAT do you need them to do?
- HOW will data help make your point?

ARTICULATE your BIG IDEA

CREATE a STORYBOARD

- → brainstorm
- → edit
- → get feedback

STEP 2: CHOOSE an APPROPRIATE VISUAL

DRAW IT!

then... CREATE it in your TOOL

iterate and look at your data different ways

SEEK FEEDBACK from others

STEP 3: ELIMINATE CLUTTER

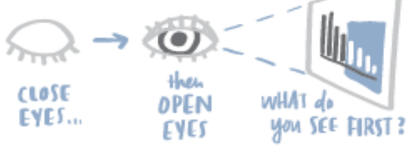
IDENTIFY UNNECESSARY ELEMENTS and REMOVE THEM

- → Leverage white space
- → Align elements
- → Avoid diagonal components

STEP 4: DRAW ATTENTION where you WANT it

Use POSITION, SIZE, and COLOR to FOCUS your AUDIENCE'S ATTENTION

Use the "where are your eyes drawn?" test



STEP 5: THINK LIKE a DESIGNER

FIRST FUNCTION → SECOND FORM

think about what you want your audience to DO with the data

create a visualization that will make this easy

- → Affordances
- → Accessibility
- → Aesthetics
- → Acceptance

STEP 6: TELL a STORY

RETURN to your STORYBOARD

Use the narrative arc to plan your story and form a pithy, repeatable phrase to help your message stick