# Final

STAT 451-01

*Jedidiah Harwood*
*Kurt Wydrinski*

*December 11, 2019*

## Contents

# 1) Color Pallets

When using color in visualizations there are different kinds of pallets that can be used.

## a. Pallet 1

Describe the pallet and explain when you would use such a pallet.

---

The palette above is an example of sequential color schemes using multi-hued colors.

Each scheme is sequential because it allows the highlighting of order in data through the use of shading. Within a given hue such as blue, there are multiple shades from light to dark. The lightness or darkness of the hue can be used to represent different levels of values. Typically, light hues represent low values and dark hues represent high values.

Each scheme is multi-hued meaning the colors used are not just shades of a single hue but instead, use multiple hues. When using multiple hues, the palettes still provide a pleasing aethetic transition from lighter shades to darker shades that preserves the implied sequential meaning of the collective colors.

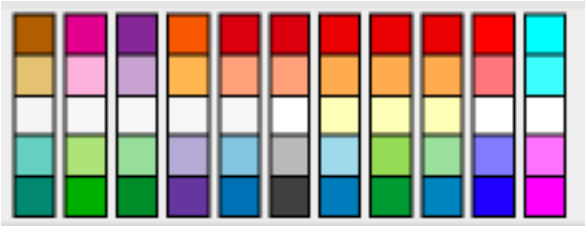
Sequential colors schemes are suited to highlighted data that can be categorized into ordered categories. Examples could be age ranges, levels of experience, density ranges, etc.

One vizualiation that may beneit from this color scheme, is a tile plot. One of the main features of a heat map, is its use of color to distinguish ordinal categorical variables, in a spacial setting. With a sequential color scheme as such, I would apply this theme to a heat map, in order to provide a well read, distinguishable, vizualization.

---

## b. Pallet 2

Describe the pallet and explain when you would use such a pallet.



---

The palette above is an example of diverging color schemes. Each scheme is diverging because it allows the highlighting of both central and extreme values in underlying data. Lighter shades and hues are used as the central colors in each of these palettes. In the palettes above, each has five (5) colors with the third (3) color

being the central, lightest color. Moving away from this color in either direction towards the first and last colors in the palette, the shades and hues get darker. The colors diverge away from a light, neutral color towards darker, more distinct colors. The colors at the ends of the palettes typically contrast highly from each other to help amplify the meaning of the divergence in the underlying data away from its central values.

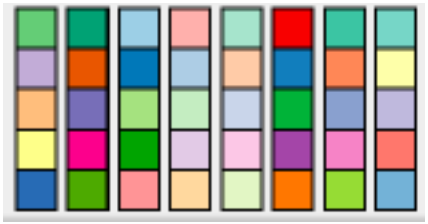
Diverging colors schemes are suited to highlight the central and extreme value in data distributions. The light coloring of central values tends to indicate the typical values of data while the bold, contrasting coloring of extreme values tends to highlight these extreme values. Examples could be grade distributions, income level distributions, age ranges, etc.

Note that many datasets could be highlighted by either sequential or diverging schemes. For example, age ranges could be highlighted by either. However, the intent of the visualization would help dictate which to use. Consider a question posed such as "Comparing pre-teen, teen, adult, and eldery populations…?" Now consider a second, similar question posed such as "What are the average ages…?" The first question is being posed from a categorial perspective that implies a sequence tied to human lifecycles. There is implied interest in order so a sequential color scheme would be applicable. For the second question, there was not much emphasis on any difference between young or old but instead more interest in the distribution, the *average*. In this case, a diverging color scheme may be more suited to the vizualization to not only highlight the average (central valeus) but also highlight the extremes.

---

## c. Pallet 3

Describe the pallet and explain when you would use such a pallet.



---

The palette above is an example of qualitative color schemes. Each scheme is designed with a set of color shades and hues that contrast from one another. Sequential and diverging color schemes do not try to contrast as much but instead try to show more relationship or transitioning of values between each color. Qualitative schemes try to show the contrast as much as possible attempting to highlighted the grouping and differences more than the similarities or nearness to other groups.

Qualitative color schemes are best used when tyring to depict different categories of data that are more distinct from each other than they are as similar or close to one another. Examples include demographic data such as racial identity, gender identity, political affiliation, religious affiliation, sports team fan affiliation, etc.

Some vizualzation to which this color scheme could be applied, include pie charts, waffle plots, donut plots, and ring plots.

---

# 2) Earthquakes

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Here is the link to the USGS website where the worldwide earthquake data can be downloaded. Download all earthquake data for the past 30 days in .csv format. Using R, make a map of the world with points where the earthquakes occurred. Make a bubble map using the magnitude. Thoroughly discuss your visualizations.

## World Map of Points

As the earthquake data of the past 30 days is updated every minute, it is important to specify that the earthquake data was downloaded at 1:08 p.m., on November 18, 2019.

To begin the map, I will read in the earthquake dataset under the variable name: "earthquake."

```
earthquake <- read_csv("earthquake.csv", col_names = TRUE)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   time = col_datetime(format = ""),
##   magType = col_character(),
##   net = col_character(),
##   id = col_character(),
##   updated = col_datetime(format = ""),
##   place = col_character(),
##   type = col_character(),
##   status = col_character(),
##   locationSource = col_character(),
##   magSource = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
head(earthquake)
```

```
## # A tibble: 6 x 22
##   time                latitude longitude  depth   mag magType   nst   gap
##   <dttm>                 <dbl>     <dbl>  <dbl> <dbl> <chr>   <dbl> <dbl>
## 1 2019-11-18 21:01:46     35.8     -119.   8.82  1.14 ml         14    88
## 2 2019-11-18 20:54:46     61.4     -150.  30.4   1.7  ml         NA    NA
## 3 2019-11-18 20:42:44     33.6     -117.  13.4   0.33 ml         17    70
## 4 2019-11-18 20:34:00     33.9     -117.  19.1   0.87 ml         25    67
## 5 2019-11-18 20:32:41     34.4     -118.  11.2   0.83 ml         12   105
## 6 2019-11-18 20:30:32     61.3     -148. 130.    1.6  ml         NA    NA
```

```
## # ... with 14 more variables: dmin <dbl>, rms <dbl>, net <chr>, id <chr>,
## #   updated <dttm>, place <chr>, type <chr>, horizontalError <dbl>,
## #   depthError <dbl>, magError <dbl>, magNst <dbl>, status <chr>,
## #   locationSource <chr>, magSource <chr>
```

Variable | Description time | Time of Earthquake occurence latitude | Latitude lLocation of Earthquake longitude | Longitude location of Earthquake depth | Depth of the Event mag | Magnitude of Event magType | Algorithm or Method Used to Evaluate the Method of the Earthquake nst | Number of Seismic Stations used to evaluate Eathquake Location gap | The Largest azimutahl gap between azimuthally adjacent stations (in degrees) horizontalError | Uncertainty of Observed Event's Location (in KM) dmin | Smallest observed Distancwe to event epicenter from the Closest Seismic Station rms | Root Mean Square Calculations of Residuals in predictions of Event occurence. net | ID of Data Contributer id | Unique Identification of Eathquake updated | Time of Upload in Original Dataset place | Nearby Named Geographical Region horizontalError | Uncertainty of Earthquake Location (in KM) depthError | Uncertainty of Earthquake Depth (in KM) magNst | Total number of Seismic Stations used to Calculate Earthquake's Magnitude Status | Indicates Whether Event has been viewed by a Person locationSource | Network that Authored location of Event magSource | Network that Authored Preffered Magnitude — | —-

---

TO DO

---

## 3) Disease / Illness Story

See `Final.pdf` in this project's Files section for detailed instructions.

---

TO DO

---