

Name: Muhammad Hafiz Wan Rosli  
Course: MAT 259: Visualizing information  
Instructor: George Legrady  
TA: Joshua Dickinson

Title: Color names in book titles

# Data Mining Assignment

## Question:

Which “colors of the rainbow” (pure spectral colors) appears most as part of a book title?

## Query:

```
SELECT
    distinct floor(deweyClass/100)*100 as dewey,

    SUM(CASE WHEN title like "violet %" or title like "% violet
    %" or title like "% violet" THEN 1 ELSE 0 END) as violet,
    SUM(CASE WHEN title like "indigo %" or title like "% indigo
    %" or title like "% indigo" THEN 1 ELSE 0 END) as indigo,
    SUM(CASE WHEN title like "blue %" or title like "% blue %"
    or title like "% blue" THEN 1 ELSE 0 END) as blue,
    SUM(CASE WHEN title like "green %" or title like "% green
    %" or title like "% green" THEN 1 ELSE 0 END) as green,
    SUM(CASE WHEN title like "yellow %" or title like "% yellow
    %" or title like "% yellow" THEN 1 ELSE 0 END) as yellow,
    SUM(CASE WHEN title like "orange %" or title like "% orange
    %" or title like "% orange" THEN 1 ELSE 0 END) as orange,
    SUM(CASE WHEN title like "red %" or title like "% red %" or
    title like "% red" THEN 1 ELSE 0 END) as red,
    SUM(CASE WHEN title like "black %" or title like "% black
    %" or title like "% black" THEN 1 ELSE 0 END) as black,
    SUM(CASE WHEN title like "white %" or title like "% white
    %" or title like "% white" THEN 1 ELSE 0 END) as white,

    SUM(CASE WHEN title like "violet %" or title like "% violet
    %" or title like "% violet" or "indigo %" or title like "%
    indigo %" or title like "% indigo" or "blue %" or title
    like "% blue %" or title like "% blue" or "green %" or
    title like "% green %" or title like "% green" or "yellow
    %" or title like "% yellow %" or title like "% yellow" or
    "orange %" or title like "% orange %" or title like "%
    orange" or "red %" or title like "% red %" or title like "%
    red" or "black %" or title like "% black %" or title like
    "% black" or "white %" or title like "% white %" or title
    like "% white" THEN 1 ELSE 0 END) as total

FROM
    transactionsall
```

```

WHERE
    deweyClass < 1000
ITEMTYPE
    acbk
GROUP BY
    floor(deweyClass/100)*100

```

- title: Curious as to which name of a color is most used in books  
Which color is most used in which deweyClass?
- deweyClass: Check color names for individual deweyClass
- group by and having: List down all color names in all dewey classes  
Group by color names

### **Processing Time:**

The process through the Sequel Pro interface on a 50/Mbps connection took 17 minutes to complete.

### **Review & Analysis:**

By doing a query on the word “red”, it is impossible to find book titles which contains the word “red”. Multiple titles that doesn’t relate to the word appears, such as: fi(red), wi(red), ti(red). By doing a three part match for this word, I am able to filter out words that has “red” in it.

“%red%” vs “% red %” vs “%red %” vs “% red%”

SUM(CASE WHEN title like "red %" or title like "% red %" or title like "% red" THEN 1 ELSE 0 END) as red

The deweyClass where color names appear the most is in the 300 class, followed closely by the 700 deweyClass, which is “Social Sciences” and “Arts” respectively. The color name that is most used is “green”.

These queries are only done for “colors of the rainbow” (pure spectral colors). These colors are represented by such names that appear on the query. However, there may be multiple representations of one particular color, such as violet vs purple.

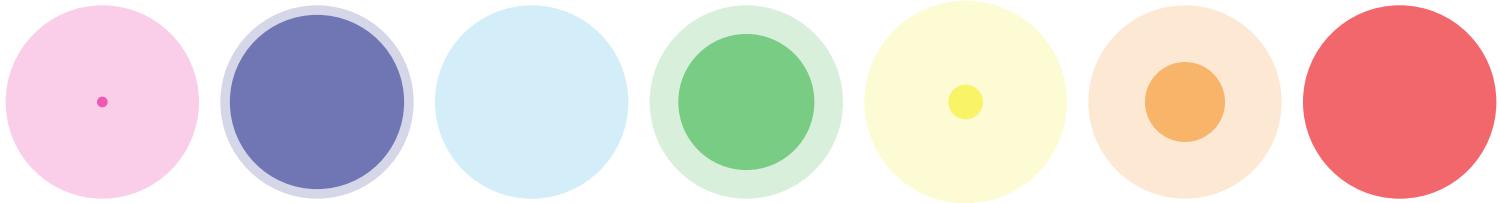
Future development on this topic should include all the possible names for each color. A more thorough research could provide a basis to question if color names could effect color psychology (as opposed to the actual color).

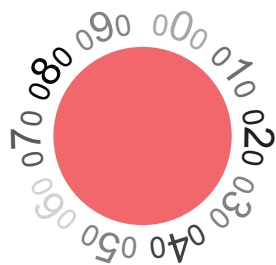
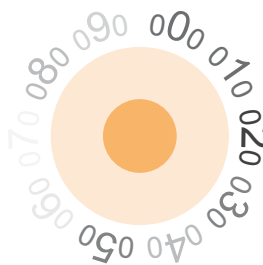
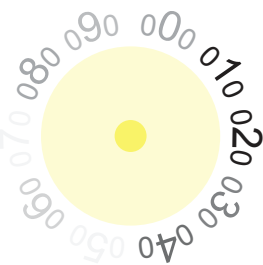
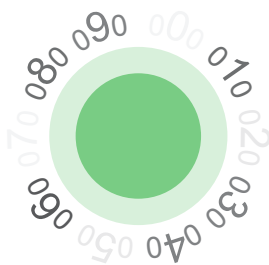
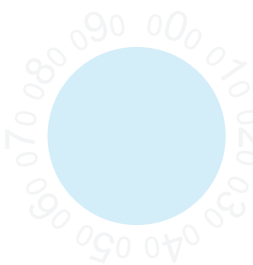
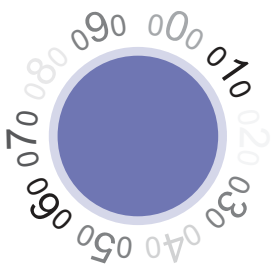
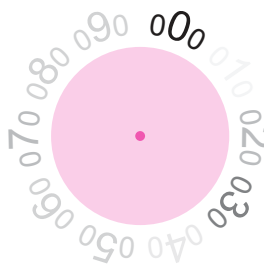
TABLES

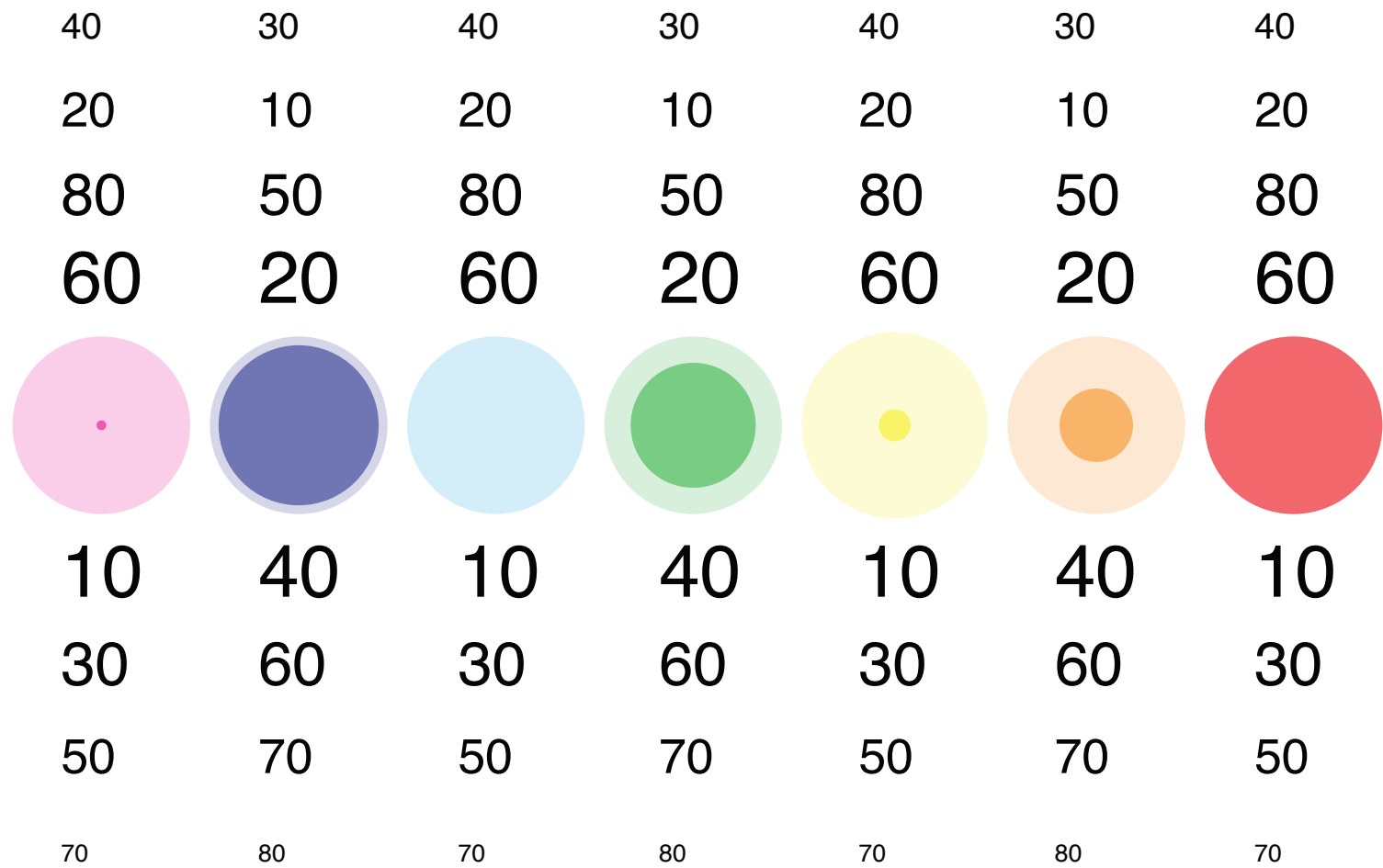
- ITEMTABLE
- tempTABLE
- trans\_2005\_thru\_2011
- transactionsall
- transactionsallID

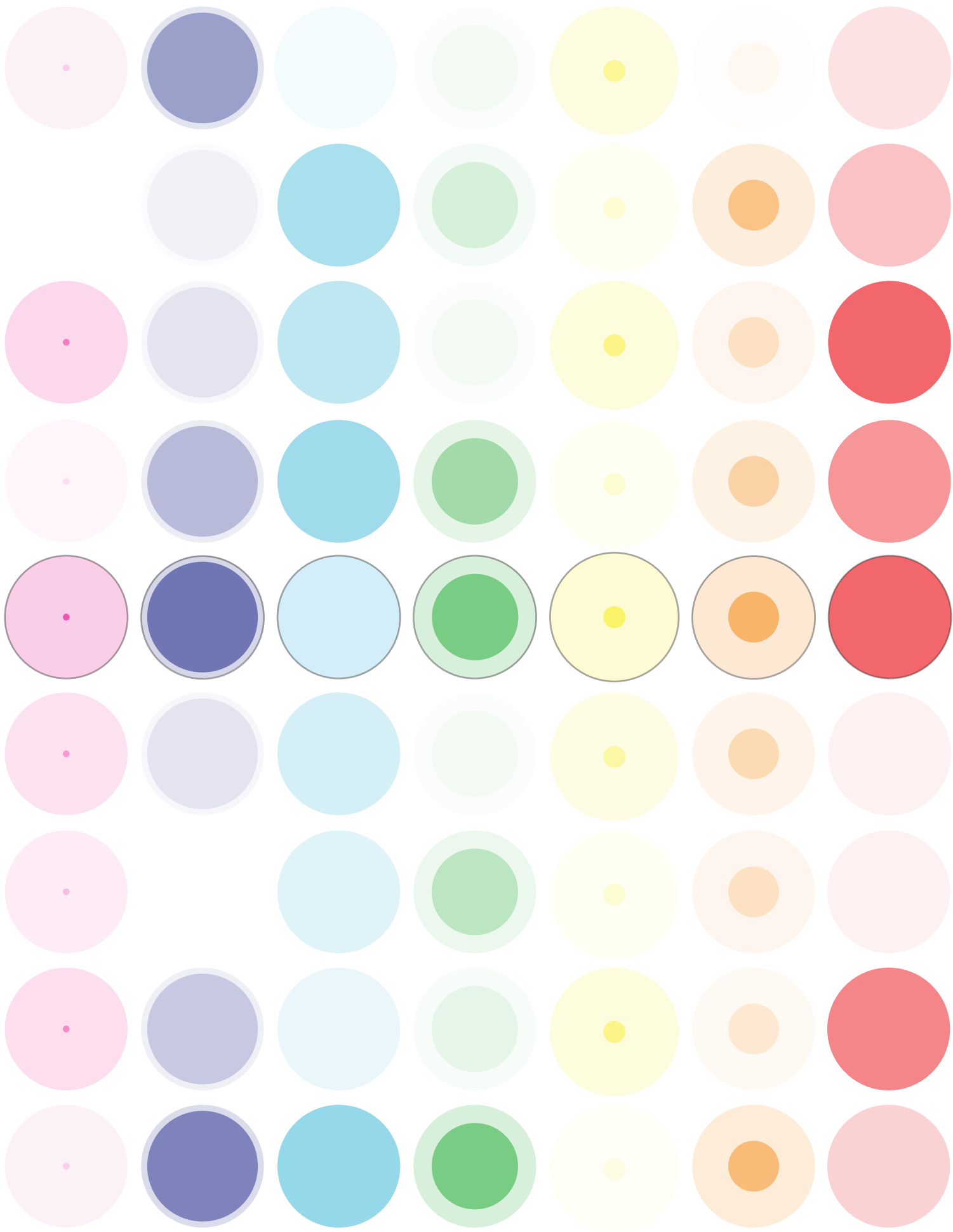
```
1 Select distinct floor(deweyClass/100)*100 as dewey, SUM(CASE WHEN title like "violet %" or title like "% violet %" or title like "% violet" THEN 1 ELSE 0 END) as violet, SUM(CASE WHEN title like "indigo %" or title like "% indigo %" or title like "% indigo" THEN 1 ELSE 0 END) as indigo, SUM(CASE WHEN title like "blue %" or title like "% blue %" or title like "% blue" THEN 1 ELSE 0 END) as blue, SUM(CASE WHEN title like "green %" or title like "% green %" or title like "% green" THEN 1 ELSE 0 END) as green, SUM(CASE WHEN title like "yellow %" or title like "% yellow %" or title like "% yellow" THEN 1 ELSE 0 END) as yellow, SUM(CASE WHEN title like "orange %" or title like "% orange %" or title like "% orange" THEN 1 ELSE 0 END) as orange, SUM(CASE WHEN title like "red %" or title like "% red %" or title like "% red" THEN 1 ELSE 0 END) as red, SUM(CASE WHEN title like "black %" or title like "% black %" or title like "% black" THEN 1 ELSE 0 END) as black, SUM(CASE WHEN title like "white %" or title like "% white %" or title like "% white" THEN 1 ELSE 0 END) as white, SUM(CASE WHEN title like "violet %" or title like "% violet %" or title like "% violet" or "indigo %" or title like "% indigo %" or title like "% indigo" or "blue %" or title like "% blue %" or title like "% blue" or "green %" or title like "% green %" or title like "% green" or "yellow %" or title like "% yellow %" or title like "% yellow" or "orange %" or title like "% orange %" or title like "% orange" or "red %" or title like "% red %" or title like "% red" or "black %" or title like "% black %" or title like "% black" or "white %" or title like "% white %" THEN 1 ELSE 0 END) as total from transactionsall where deweyClass < 1000 and itemtype = "acbk" group by floor(deweyClass/100)*100 with rollup;
```

Query Favorites Query History Run Current Run All											
dewey	violet	indigo	blue	green	yellow	orange	red	black	white	total	
0	0	0	203	16	32	0	713	3051	250	1622	
100	0	403	272	512	0	0	489	761	174	1347	
200	10	0	1118	416	0	458	593	829	424	2157	
300	88	0	4650	10659	474	592	3673	13487	12820	32146	
400	0	0	64	0	30	0	395	94	349	691	
500	0	0	986	562	0	0	934	1833	625	3171	
600	0	242	4129	16623	593	367	2948	5045	2034	20401	
700	177	274	5267	11896	1337	809	5336	25025	8567	31182	
800	65	23	1668	607	270	46	1633	3790	2076	5657	
900	127	0	2669	2389	298	161	4265	8441	9186	18166	
NULL	467	942	21026	43680	3034	2433	20979	62356	36505	116540	





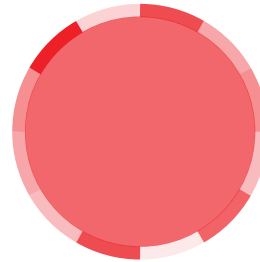
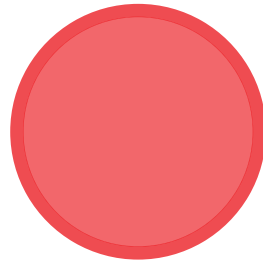
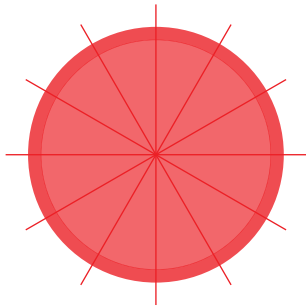




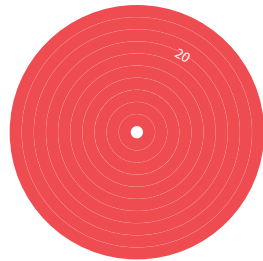
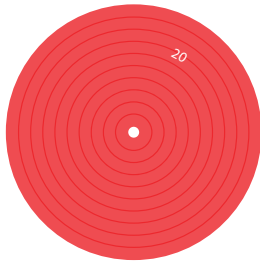




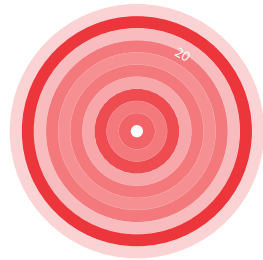
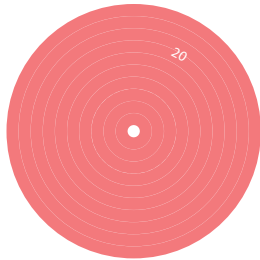
## Studies: divisions of a circle



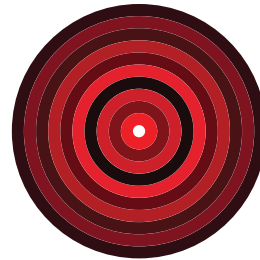
donut chart = central angle arc length



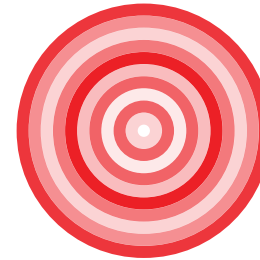
radial bar



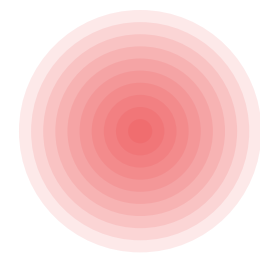
saturation



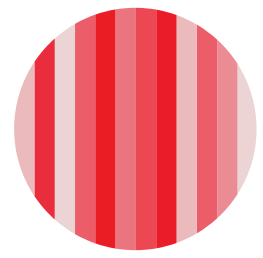
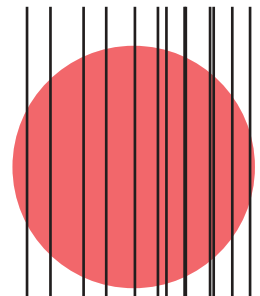
brightness



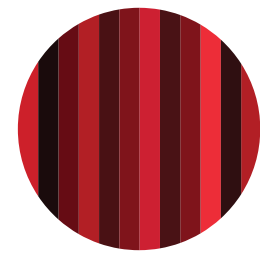
opacity



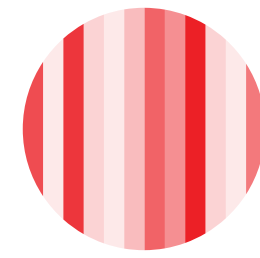
opacity 2



saturation



brightness



opacity