

# EKSPLORASI DATA PENGELOMPOKKAN BINTANG NASA : FAKTOR DAN KETERKAITAN ANTAR VARIABEL

Intan Citra Phonskaningtyas (06211940000007)

Jovanka Alvira Wijaya(06211940000079)

Departemen Statistika ITS/ Analisis Eksplorasi Data C

Institut Teknologi Sepuluh Nopember

Email : intanc110@gmail.com, jovankaawijaya10@gmail.com

## I. PENDAHULUAN

### 1.1 Latar belakang

Bintang terbentuk dari kumpulan awan dan debu angkasa yang bersatu selama lebih dari satu tahun cahaya. Gravitasi adalah faktor yang menarik debu-debu angkasa untuk bersatu dan menjadi bintang dengan bentuk bola yang sangat besar, hasil reaksi tersebut membuat inti bintang begitu panas hingga dapat mencapai 15 juta derajat. Bintang baru memiliki variasi ukuran dan warna, dengan jangkauan warna biru hingga merah berdasarkan suhu yang dimiliki bintang tersebut. NASA (*National Aeronautics and Space Administration*) mengklasifikasikan bintang yang ada di alam semesta menjadi beberapa tipe, 5 diantaranya yakni *red dwarf*, *brown dwarf*, *white dwarf*, *main sequence*, *super giants*, dan *hyper giants*.

*Red dwarf* adalah tipe bintang yang memiliki temperatur dan ukuran terkecil jika dibandingkan dengan bintang tipe lainnya.

Saat bintang ini mati karena kehabisan bahan bakar yang dihasilkan oleh bintang itu sendiri, *red dwarf* akan berubah menjadi bintang tipe *white dwarf*. *Brown dwarf* biasa disebut oleh para saintis dengan *the failed stars* atau bintang gagal, adalah bintang yang memiliki ukuran kecil sehingga tidak seperti bintang lainnya bintang ini terkadang tidak memiliki cukup fusi hidrogen. *White dwarf* adalah bintang yang telah kehabisan bahan nuklir dari inti bintang sehingga *white dwarf* dapat dikatakan sebagai jenis yang muncul ketika bintang lainnya mulai kehabisan bahan bakar atau dalam tahapan menjadi bintang mati. *Main sequence stars* adalah tipe bintang yang paling banyak ditemukan saat ini, disebut *main sequence* karena lokasi spesifik pusat massa bintang terletak pada *main sequence* nya. *Super giants* dan *hyper giants* secara garis besar dibedakan melalui ukuran dan umur bintang, dimana *hyper giants* akan memiliki umur dan ukuran yang lebih besar dibandingkan *supergiants*. Ada banyak faktor lain yang disebutkan memiliki

andil dalam pengelompokkan bintang-bintang menurut NASA seperti radius, *magnitude*, warna, dan lainnya. Salah satu tujuan dibentuknya pengklasifikasi bintang adalah agar manusia dapat membedakan ciri bintang satu dan lainnya, selain itu untuk menunjang kemudahan para saintis untuk melakukan penelitian lebih lanjut.

Dalam karya kali ini, akan dilakukan eksplorasi data mengenai pengaruh faktor-faktor yang digunakan sebagai pengklasifikasi tipe bintang dan hubungan apa yang dimiliki antar faktor tersebut, karena hubungan ini sangat penting untuk mengetahui apakah ada keterkaitan antar faktor dan bagaimana hal tersebut dapat membuat suatu bintang diklasifikasikan dalam tipe tertentu.

### 1.2 Rumusan masalah

Rumusan masalah dalam karya kali ini adalah bagaimana eksplorasi data dari faktor-faktor pengklasifikasi tipe bintang

dan bagaimana hubungan yang dimiliki antar variabelnya?

### 1.3 Tujuan

Tujuan dibuatnya karya ini adalah untuk mengetahui eksplorasi data dari faktor-faktor pengklasifikasi tipe bintang dan bagaimana hubungan yang dimiliki antar variabelnya

## II. METODOLOGI

### 2.1. Sumber data

Data yang digunakan dalam karya kali ini adalah data Star Type Classification/NASA oleh Baris Dincer yang diambil dari platform repository data Kaggle. Data yang digunakan dilakukan pengambilan pada:

Hari/Tanggal : Rabu, 23 Juni 2021

Pukul : 08.00 WIB

### 2.2. Variabel Penelitian

Variabel yang digunakan berjumlah 7 variabel dengan keterangan sebagai berikut

Variabel	Tipe Data	Deskripsi
<i>Star Type</i>	Nominal	Tipe bintang (Red Dwarfs - 0, Brown Dwarfs - 1, White Dwarfs - 2, Main Sequence - 3, Super Giants - 4, Hyper Giants - 5)
<i>Temperature</i>	Numerik	Suhu bintang (1939°K - 40000°K)
<i>Relative Luminosity</i>	Numerik	Total energi yang dipancarkan suatu bintang ke segala arah per detik (8e-05 L <sub>☉</sub> - 849420 L <sub>☉</sub> )
<i>Relative Radius</i>	Numerik	Radius bintang (0.0084 R <sub>☉</sub> - 1984.5 R <sub>☉</sub> )
<i>Absolute Magnitude</i>	Numerik	Kecerahan bintang dengan keterangan semakin kecil magnitude, semakin terang bintang tersebut (-11.92 Mw - 20.06 Mw)

Variabel	Tipe Data	Deskripsi
<i>Color</i>	Nominal	Warna bintang, semakin merah suatu bintang maka suhunya semakin tinggi (17 Unique Values)
<i>Spectral Class</i>	Nominal	Klasifikasi kelas spektrum bintang (M,B,A,F,O,K,G)

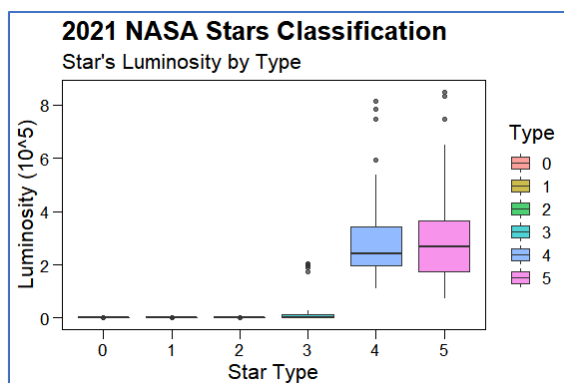
### 2.3. Langkah Analisis

Analisis yang dilakukan dalam eksplorasi data kali ini adalah sebagai berikut.

- Merumuskan masalah
- Melakukan studi literatur
- Mencari data yang sesuai
- Melakukan eksplorasi data dengan terlebih dahulu membersihkan data dan menyiapkan data sebelum dilakukannya eksplorasi lebih lanjut dengan *software R*
- Melakukan analisis dari hasil eksplorasi yang dilakukan
- Melakukan penarikan kesimpulan

## III. ANALISIS

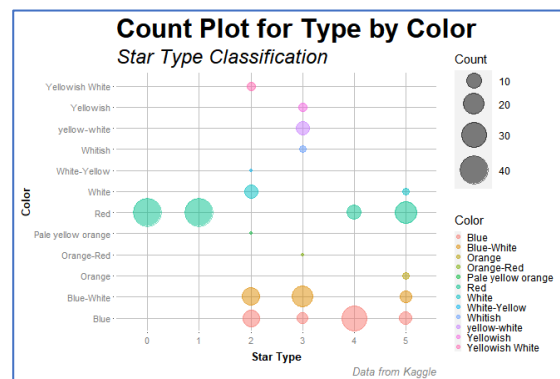
### 3.1. Perbandingan tipe bintang dengan variabel lain



**Gambar 3.1.** Stars Luminosity Box Plot

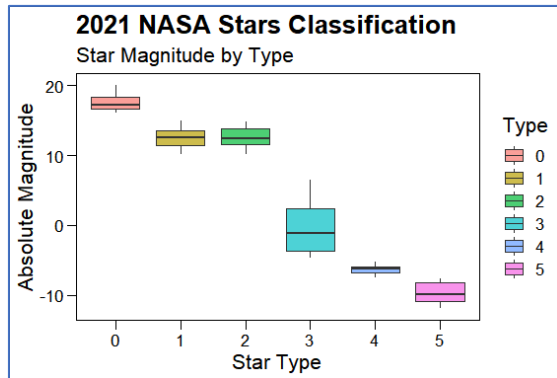
Luminositas adalah energi yang dipancarkan oleh bintang dari segala arah dalam setiap

detiknya. Jika dibandingkan, tipe bintang 0-3 tidak memiliki perbedaan luminositas yang signifikan dan nilai energi yang dipancarkan lebih kecil jika dibandingkan dengan bintang tipe 4 dan 5. Pada tipe bintang 3, 4, dan 5 memiliki beberapa *outlier* yang cukup jauh, serta persebaran data luminositas bintang 4 dan 5 tidak terlihat adanya perbedaan.



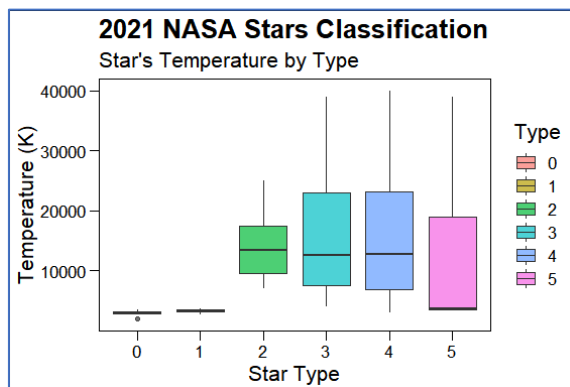
**Gambar 3.2.** Stars Color Count Plot

Gambar di atas merupakan *countplot* tipe bintang berdasarkan warna. Dapat dilihat bahwa tipe bintang 0 dan 1 hanya merupakan bintang berwarna merah, tipe bintang 2, 3, dan 5 memiliki jumlah bintang berwarna biru dan biru muda lebih banyak dibanding warna lainnya.



**Gambar 3.3.** *Star Magnitude Box Plot*

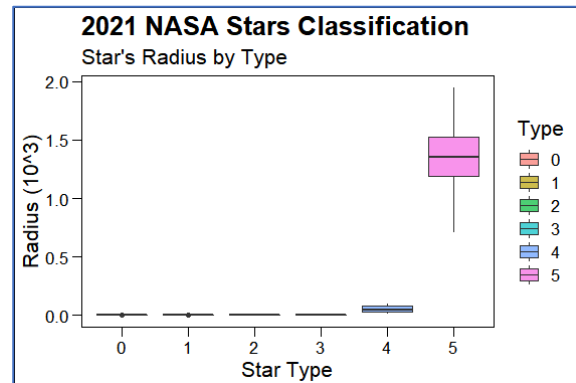
Magnitude atau dalam hal ini adalah kecerahan bintang jika ditampilkan dalam visualisasi *boxplot* seperti di atas, tipe bintang dengan kode terkecil memiliki kecerahan yang lebih besar atau bintang yang paling terang dibandingkan tipe lainnya. Berdasarkan visualisasi di atas, semakin bertambah kode tipe bintang, semakin redup bintang tersebut.



**Gambar 3.4.** *Stars Temperature Box Plot*

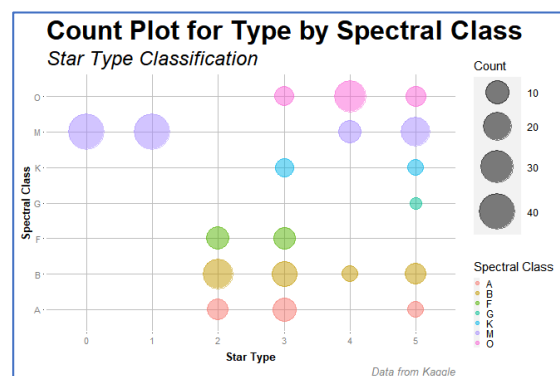
Berdasarkan visualisasi suhu berdasarkan tipe bintang, tidak ada perbedaan antara tipe bintang 0 dan 1, begitu pula dengan tipe bintang 2, 3, 4, dan 5 walaupun memiliki temperatur yang lebih tinggi dibandingkan tipe 0 dan 1 namun antar tipe bintang

memiliki persebaran yang merata sehingga tidak dapat ditarik kesimpulan bahwa temperatur mempengaruhi tipe bintang.



**Gambar 3.5.** *Stars Radius BoxPlot*

Radius bintang adalah jarak dari pusat bintang ke bagian terluar bintang. Dalam visualisasi perbandingan radius bintang dengan tipe bintang, tidak ada perbedaan radius yang signifikan antara tipe bintang 0-4, namun pada tipe bintang 5 radius bintang sangat besar karena sesuai dengan tipenya bintang ini disebut sebagai *hyper-giant stars* atau bintang yang sangat besar.

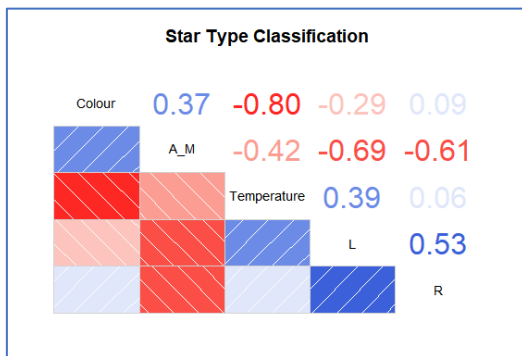


**Gambar 3.6.** *Stars Spectral Class Count Plot*

Tidak ada pola yang signifikan dapat terlihat dari pengelompokkan tipe berdasarkan kelas

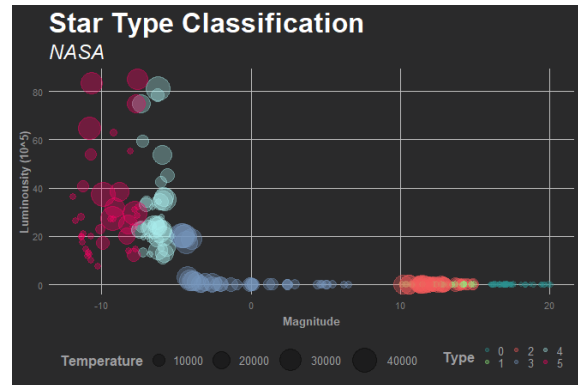
spektral, hanya saja jika dilihat lebih detail maka beberapa tipe bintang yang berdekatan memiliki bintang dengan kelas spektral yang sama. Contoh-nya tipe 0 dan 1, keduanya hanya memiliki bintang dengan *spectral class* M, lalu tipe 2 dan 3 yang memiliki bintang dengan kelas spektral F, B, dan A, serta tipe 3, 4, dan 5 yang memiliki bintang dengan spektral kelas B dan O.

### 3.2. Hubungan antar Variabel



**Gambar 3.7.** Correlation Plot

Dari plot korelasi di atas dapat diketahui beberapa hal, yakni korelasi antar variabel terbesar ada pada variabel warna bintang dengan temperatur, warna bintang diurutkan dari biru hingga merah yang menandakan semakin tinggi temperatur maka semakin biru warna bintang yang dihasilkan. Variabel lain yang terlihat memiliki korelasi yang besar adalah *magnitude* dengan luminositas, serta luminositas dengan radius. Keterangan warna pada *correlation plot* yaitu merah untuk korelasi negatif dan biru untuk korelasi positif.



**Gambar 3.9.** Bubble Plot antar variabel

Plot di atas merupakan gambaran dari perbandingan yang dilakukan antar variabel tipe bintang, *absolute magnitude*, Luminositas, dan temperatur. Jika dilihat berdasarkan tipe bintang, tipe bintang 5 memiliki magnitude yang lebih rendah jika dibandingkan dengan tipe bintang lain dan temperatur bintang bertambah seiring dengan bertambahnya luminositas. Bintang dengan tipe 4 memiliki ciri yang sama dengan tipe 5 hanya saja temperatur yang dimiliki lebih tinggi. Dapat dilihat pada visualisasi di atas bahwa kode tipe bintang akan bertambah seiring dengan turunnya temperatur, lalu pada tipe 0 hingga 2 memiliki luminositas yang sangat kecil dan temperaturnya akan bertambah seiring berkurangnya magnitude bintang.

## IV. PENUTUP

### 4.1. Kesimpulan

Berdasarkan hasil analisis di atas, dapat ditarik beberapa kesimpulan yaitu :

- Dalam hal klasifikasi bintang menjadi 5 tipe yang telah disebutkan, luminositas dan radius tidak terlalu berpengaruh dalam penentuan pengelompokkan variabel kecuali pada tipe 5 yang memiliki radius jauh lebih besar.
- Dari beberapa variabel, *magnitude* bintang adalah salah satu variabel yang terlihat jelas pengaruhnya terhadap penentuan tipe bintang jika dilihat melalui visualisasi. Dapat dilihat bahwa semakin besar *magnitude* bintang, semakin tinggi pula kode tipe bintang.
- Hubungan antar variabel faktor pengelompok tipe bintang mempunyai hubungan yang cukup beragam. Hubungan tertinggi terdapat pada variabel warna dan temperatur, serta *magnitude* dan luminositas. Sedangkan yang tidak signifikan hubungannya adalah warna bintang dan temperatur jika masing-masing dibandingkan dengan radius bintang. Hubungan antar variabel juga dapat dibedakan lagi jika

dilihat karakteristik setiap tipe bintang yang berbeda.

### DAFTAR PUSTAKA

Dincer, Baris. *Star Type Classification* NASA.<https://www.kaggle.com/brsdincer/star-type-classification> (Diakses pada 23 Juni 2021).

Granger, Kara C. & Whitlock, Laura A., 1998. *The Life Cycle of Stars*. United States of America: NASA.

R Cookbook. *Renaming levels of a factor*. ([http://www.cookbook-r.com/Manipulating\\_data/Renaming\\_levels\\_of\\_a\\_factor/](http://www.cookbook-r.com/Manipulating_data/Renaming_levels_of_a_factor/), diakses pada 27 Juni 2021).

Wikipedia. *Asteroid Spectral Types*. (Online). ([https://en.wikipedia.org/wiki/Asteroid\\_spectral\\_types](https://en.wikipedia.org/wiki/Asteroid_spectral_types), diakses pada 26 Juni 2021).

Stackoverflow. 2016. *bubble chart: how to change names and range of a tile, and the layout of the plot* (<https://stackoverflow.com/questions/37891590/bubble-chart-how-to-change-names-and-range-of-a-tile-and-the-layout-of-the-plo>, diakses pada 26 Juni 2021)

## LAMPIRAN I: DATA

Data yang digunakan berupa 7 kolom dan 241 baris

Sumber data : <https://www.kaggle.com/brsdincer/star-type-classification>

Temperature	L	R	A_M	Color	Spectral_Class	Type
3068	0.0024	0.17	16.12	Red	M	0
3042	0.0005	0.1542	16.6	Red	M	0
2600	0.0003	0.102	18.7	Red	M	0
2800	0.0002	0.16	16.65	Red	M	0
1939	0.000138	0.103	20.06	Red	M	0
2840	0.00065	0.11	16.98	Red	M	0
2637	0.00073	0.127	17.22	Red	M	0
2600	0.0004	0.096	17.4	Red	M	0
2650	0.00069	0.11	17.45	Red	M	0
2700	0.00018	0.13	16.05	Red	M	0
3600	0.0029	0.51	10.69	Red	M	1
3129	0.0122	0.3761	11.79	Red	M	1
3134	0.0004	0.196	13.21	Red	M	1
3628	0.0055	0.393	10.48	Red	M	1
2650	0.0006	0.14	11.782	Red	M	1
3340	0.0038	0.24	13.07	Red	M	1
2799	0.0018	0.16	14.79	Red	M	1
...	...	...	...	...	...	...
9383	342940	98	-6.98	Blue	O	4
23440	537430	81	-5.975	Blue	O	4
16787	246730	62	-6.35	Blue	O	4
18734	224780	46	-7.45	Blue	O	4
9892	593900	80	-7.262	Blue	O	4
10930	783930	25	-6.224	Blue	O	4
23095	347820	86	-5.905	Blue	O	4
21738	748890	92	-7.346	Blue	O	4
24145	382993	1494	-8.84	Blue-white	B	5
38234	272830	1356	-9.29	Blue	O	5
32489	648430	1948.5	-10.84	Blue	O	5
27739	849420	1252	-7.59	Blue-white	B	5
21904	748490	1130	-7.67	Blue-white	B	5
38940	374830	1356	-9.93	Blue	O	5
30839	834042	1194	-10.63	Blue	O	5
8829	537493	1423	-10.73	White	A	5
9235	404940	1112	-11.23	White	A	5
37882	294903	1783	-7.8	Blue	O	5

## LAMPIRAN II: SYNTAX R

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(corrgram)
```

```
library(ggthemes)
```

```
library(plyr)
```

```
library(RColorBrewer)
```

```
#Import Data
```

```
df3 = read.table("E:/AED/Stars.csv",sep=',',header=T)
```

```
#Buat Tema
```

```
theme_star = function(){
```

```
  theme(plot.title = element_text(size=24, face='bold'),
        plot.subtitle = element_text(size=18, face='italic'),
        plot.caption = element_text(size=10, face='italic', colour='grey50'),
        axis.title.x = element_text(size = 10, face = 'bold', vjust = 0),
        axis.text.x = element_text(size = 8, color= 'grey50'),
        axis.title.y = element_text(size = 10, face = 'bold', vjust = 1),
        axis.text.y = element_text(size = 8, color= 'grey50'),
        axis.ticks = element_line(linetype=3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.grid.major.y = element_line(colour = 'grey75'),
        panel.grid.major.x = element_line(colour = 'grey75'),
        panel.border = element_blank(),
        panel.background = element_blank())
```

```
}
```

```
#Cleaning Data
```



```

data = df3 %>%
  within({
    Spectral_Class=factor(Spectral_Class)
    Type=factor(Type)})

df = df3$Color
df= mapvalues(df, from = c("Blue white", "Blue White", "Blue-white", "white", "yellowish"
), to = c('Blue-White','Blue-White','Blue-White','White','Yellowish'))
df = as.factor(df)

df=data.frame(df)

data2=cbind(data,df)

data$Color= mapvalues(data$Color, from = c("Blue white", "Blue White", "Blue-white",
"white", "yellowish" ), to = c('Blue-White','Blue-White','Blue-White','White','Yellowish'))
data$Color= mapvalues(data$Color, from = c("Red", "Blue-White", "White", "Yellowish",
"Whitish", "Pale yellow orange"), to = c("12", "2", "4", "7", "3", "9"))
data$Color= mapvalues(data$Color, from = c("Yellowish White", "yellow-white", "Blue",
"Orange"), to = c("5", "6", "1", "10"))
data$Color= mapvalues(data$Color, from = c("White-Yellow", "Orange-Red"), to =
c("8", "11"))

data_2= data %>%
  mutate(Colour=as.numeric(Color))

#Korelasi Keseluruhan Faktor
corrgram(data_2 %>% select(where(is.numeric)), order=TRUE,
  upper.panel=panel.cor, main="Star Type Classification")

#Perbandingan Tiap variabel Numerik dengan Star Type
data_2 %>%

```

```
ggplot(aes(x=Type, y=A_M, fill=Type)) + geom_boxplot(alpha=0.7) +
labs(title="2021 NASA Stars Classification",
      subtitle="Star Magnitude by Type",
      x = "Star Type",
      y = "Absolute Magnitude") + theme_star() +
theme_base()
```

data\_2 %>%

```
ggplot(aes(x=Type, y=Temperature, fill=Type)) + geom_boxplot(alpha=0.7) +
labs(title="2021 NASA Stars Classification",
      subtitle="Star's Temperature by Type",
      x = "Star Type",
      y = "Temperature (K)") + theme_star() +
theme_base()
```

data\_2 %>%

```
ggplot(aes(x=Type, y=L/100000, fill=Type)) + geom_boxplot(alpha=0.7) +
labs(title="2021 NASA Stars Classification",
      subtitle="Star's Luminosity by Type",
      x = "Star Type",
      y = "Luminosity (10^5)") + theme_star() +
theme_base()
```

data\_2 %>%

```
ggplot(aes(x=Type, y=R/1000, fill=Type)) + geom_boxplot(alpha=0.7) +
labs(title="2021 NASA Stars Classification",
      subtitle="Star's Radius by Type",
      x = "Star Type",
      y = "Radius (10^3)") + theme_star() +
theme_base()
```

#Perbandingan Variabel Nominal dengan Star Type

```
data_2 %>%
  ggplot(aes(Type, Spectral_Class, color=Spectral_Class)) + geom_count(alpha=0.5) +
  scale_size(range=c(5,15)) + theme_star() +
  labs(title="Count Plot for Type by Spectral Class",
        subtitle="Star Type Classification",
        caption="Data from Kaggle",
        x="Star Type",
        y="Spectral Class",
        color="Spectral Class",
        size="Count") +
  theme(legend.key.size = unit(0.2, 'cm'))
```

```
data2 %>%
  ggplot(aes(Type, df, color=df)) + geom_count(alpha=0.5) +
  scale_size(range=c(1,12)) + theme_star() +
  labs(title="Count Plot for Type by Color",
        subtitle="Star Type Classification",
        caption="Data from Kaggle",
        x="Star Type",
        y="Color",
        color="Color",
        size="Count") +
  theme(legend.key.size = unit(0.2, 'cm'))
```

#Bubble Plot

```
data_2 %>%
  ggplot(aes(x = A_M, y = L/10000, size = Temperature, color=Type)) +
  geom_point(alpha = 0.33, position = 'jitter') +
  labs(title= "Star Type Classification",
        subtitle= "NASA",
        x = 'Magnitude',
```

```
y = 'Luminosity (10^5)' +  
scale_size(range = c(.5, 10)) +  
theme(legend.key.size = unit(0.3, 'cm'), legend.title = element_text(face='bold')) +  
theme_hc(style='darkunica') + scale_colour_hc('darkunica') + theme_star()
```