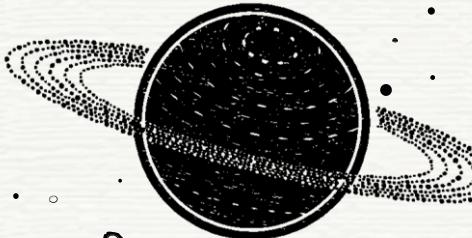
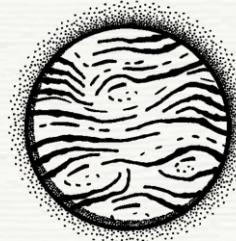


# ANALISI DEL DATASET:

# CLASSIFICAZIONE DELLE STELLE

---

Erika Amato  
Carlotta Bellomo  
Riccardo Mirarchi  
Annalisa Terzoli



# **INTRODUZIONE**

Abbiamo deciso di analizzare un dataset riguardante la classificazione di alcune stelle fornita dalla NASA.

Il dataset è stato analizzato sia da un punto di vista **teorico** che **pratico**.

Abbiamo cercato di applicare le norme teoriche per arrivare ad una buona rappresentazione dei dati.

Nella prima parte della presentazione sono proposte le **regole teoriche** che abbiamo cercato di applicare nella successiva fase grafica.

La seconda parte è dedicata all'**analisi del dataset**. Abbiamo approfondito la conoscenza delle proprietà astronomiche presenti nel dataset in modo da poter comprendere meglio i dati in nostro possesso.

- In questa presentazione vengono proposti i **grafici statici**. In aggiunta, sui notebook disponibili al [link GitHub](#), sono presenti i **grafici interattivi** che abbiamo elaborato e che possono aiutare a comprendere meglio l'analisi del dataset effettuata.

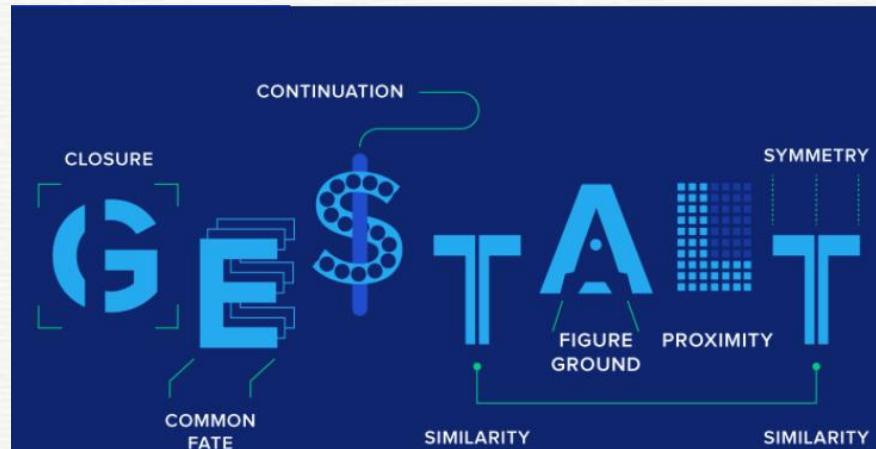


# COME VISUALIZZARE I DATI: UNA BREVE SPIEGAZIONE

---

# PRINCIPI FONDAMENTALI DEL DESIGN

I principi della Gestalt sono un'importante insieme di idee e la loro implementazione oltre a migliorare l'estetica del design ne migliora la funzionalità. Più semplicemente la teoria su cui sono basati i principi è fondata sulla funzionalità del cervello umano, su come tenda a semplificare e percepire immagini complicate composte da molti elementi. Il cervello, infatti, cerca di trovare strutture e patterns per comprendere meglio l'ambiente in cui viviamo.



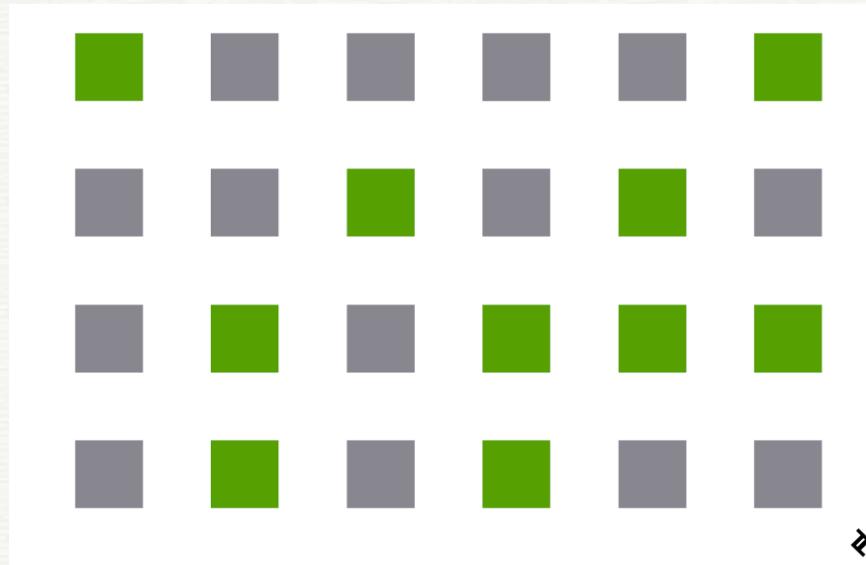
# **PRINCIPI FONDAMENTALI DEL DESIGN**

I principi sono sei:

- Somiglianza
  - Continuità
  - Chiusura
  - Prossimità
  - Figure/ground
  - Simmetria e ordine
- A questi sei si possono aggiungere altri principi fra cui quello chiamato ‘Common fate’

# **PRINCIPIO DI SOMIGLIANZA**

Somiglianza tra colori e forme permettono di raggruppare gli elementi per categorie, anche se non si trovano visivamente vicine.



# <sup>+</sup>PRINCIPIO DI CONTINUITÀ

Per la legge di continuità l'occhio umano tende a seguire il percorso più lineare indipendentemente da come le linee sono disegnate



# <sup>+</sup>PRINCIPIO DI CHIUSURA

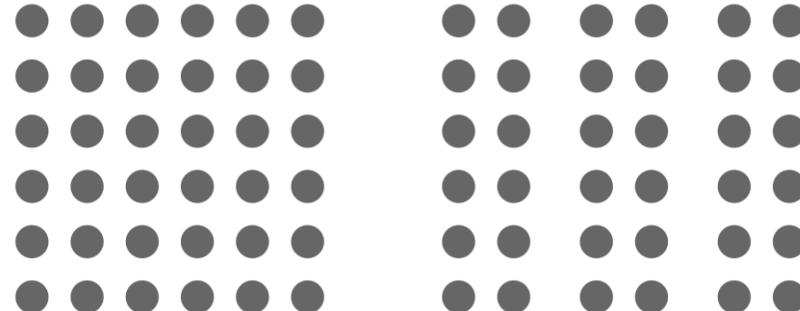
Il principio di chiusura si basa sull'idea che il cervello è in grado di “chiudere” le parti mancanti dell'immagine per ricostruirla



# <sup>+</sup>PRINCIPIO DI PROSSIMITÀ

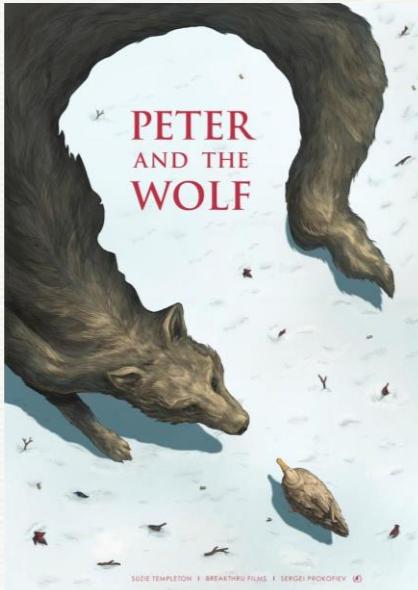
La vicinanza tra elementi permette la definizione di raggruppamenti, o meglio seconda della distanza tra gli oggetti il cervello umano identifica gli elementi come suddivisi in uno o più gruppi.

Differenze di posizione all'interno di un grafico permettono di identificare valori anomali.



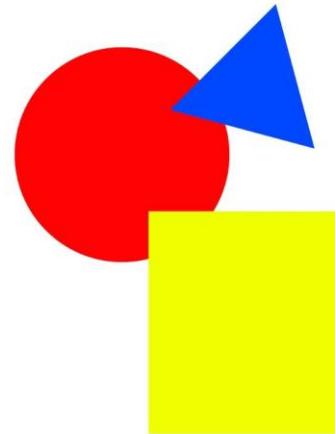
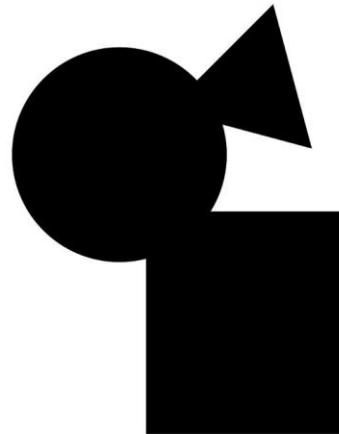
# PRINCIPIO DI FIGURE/GROUND

Simile al principio di chiusura è quello figure/ground. Il cervello viene ingannato dalla presenza di più elementi all'interno dell'immagine: l'area più grande dell'immagine viene interpretata come sfondo mentre la più piccola come figura.



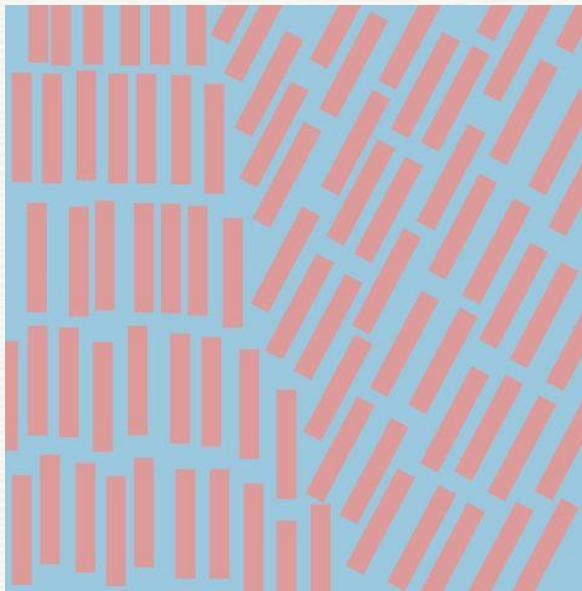
# **PRINCIPIO DI SIMMETRIA E ORDINE**

La legge sulla simmetria e l'ordine, definita in tedesco prägnanz che significa “buona figura”, afferma che il cervello cerca di ricondurre gli elementi dell'immagine a forme note



# <sup>+</sup> PRINCIPIO ‘COMMON FATE’

Il common fate afferma che visivamente si tende a raggruppare oggetti che puntano o si muovono nella stessa direzione.



# **INK RATIO**

Per data-ink ratio intendiamo la proporzione di inchiostro che viene usata per presentare i dati comparato alla quantità totale di inchiostro o pixel utilizzati. Per questo motivo una **buona rappresentazione** è quella che contiene solo l'**inchiostro necessario** per rappresentare i dati.

Nel digitale non si fa riferimento all'inchiostro ma ci si riferisce alle seguenti formule:

- **Data-ink:** rappresenta il minimo di elementi di un diagramma necessari per rappresentare un set di dati
  - **Total ink:** rappresenta tutti gli elementi usati per creare il diagramma
- Tufte, pioniere della visualizzazione dei dati, ci indica cinque regole che andrebbero seguite nella rappresentazione dei grafici:
1. Mostra soprattutto i dati
  2. Massimizza il data-ink ratio
  3. Cancella tutto ciò che non è dati
  4. Cancella i dati ridondanti
  5. Rivedi e modifica

# ATTRIBUTI PREATTENTIVI: IL COLORE

Quando si lavora col colore nei grafici è necessario seguire delle regole per poter rappresentare correttamente i dati.

1. **Usa il colore consapevolmente:** è consigliabile scegliere in prima persona il colore per raggiungere lo scopo comunicativo desiderato e non lasciare la scelta in mano al software
2. **Usare il colore con parsimonia:** non è necessario utilizzare tanti colori perché si rischia di confondere l'osservatore
3. **Attenzione alla coerenza:** se si sceglie un colore per un dato è buona norma mantenerlo per tutti i dati dello stesso tipo che verranno rappresentati
4. **Uso delle scale cromatiche:** l'intensità del colore è direttamente collegata con la progressione del dato
5. **Impatto emotivo del colore:** ogni colore viene associato inconsciamente ad un'emozione, per questo è necessario scegliere con parsimonia i colori

# I TIPI DI DATO

Esistono due tipi di dati generici:

- **Quantitativi:** le variabili statistiche quantitative sono variabili quantificabili, discrete o continue. Sono discrete se sono quantificabili con l'insieme dei numeri interi. Sono continue se sono quantificabili con l'insieme dei numeri reali. Una variabile quantitativo può essere
  - **Scalare:** le variabili statistiche sono scalari quando è possibile calcolare la differenza tra due valori, quindi anche di ordinarli. Si misurano così le variabili quantitative per le quali lo zero è convenzionale (arbitrario). In tal caso non ha senso rapportare le misure ottenute, ed è invece corretto confrontare per differenze.
  - **Proporzionale:** le variabili statistiche sono proporzionali quando è possibile calcolare la proporzione tra due valori. Si misurano così le variabili quantitative per le quali lo zero è naturale

# I TIPI DI DATO

- **Qualitativi:** le variabili statistiche qualitative identificano una particolare qualità (modalità) e non sono quantificabili. Sono dette anche mutabili statistiche. Un carattere qualitativo può essere ordinale o nominale.
- **Ordinale:** Le variabili statistiche sono ordinabili quando è possibile definire un ordine delle modalità tramite un criterio oggettivo. Se una variabile è misurata su scala ordinale, si possono instaurare le seguenti relazioni tra le modalità:

$$x_i \leq x_j \text{ oppure } x_i \geq x_j$$

- **Nominale:** le variabili statistiche sono sconnesse quando le modalità, quantitative o qualitative, non sono ordinabili mediante un criterio oggettivo (variabili statistiche non ordinabili). Se una variabile è misurata su scala nominale, si possono instaurare solo le seguenti relazioni tra le modalità:

$$x_i = x_j \text{ oppure } x_i \neq x_j$$

# NORMALIZZAZIONE

In statistica, la **normalizzazione** consiste essenzialmente nel limitare l'escursione di un insieme di valori entro un certo intervallo predefinito.

Nella normalizzazione, i dati vengono **ridimensionati su un intervallo fisso**, in genere da **0 a 1**. E' una buona tecnica da utilizzare quando non si conosce la distribuzione dei dati o quando si sa che la distribuzione non è gaussiana (ossia non ha la forma di una curva a campana).

Formula per normalizzare i dati:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

# NORMALIZZAZIONE

La normalizzazione rende i dati più adatti ad una **convergenza** e una **comparazione**.

Lo svantaggio principale della normalizzazione min-max è che essa non gestisce molto bene i valori anomali. Ad esempio, se si dispone di 99 valori tra 0 e 40 e un valore è pari a 100, i 99 valori verranno tutti trasformati in un valore compreso tra 0 e 0,4.

La normalizzazione min-max garantisce che tutte le caratteristiche abbiano la stessa identica scala ma non gestiscono correttamente i valori anomali.

# BARPLOT O BARCHART

Un grafico a barre mostra i dati generali come **barre rettangolari** con altezze proporzionali al valore che rappresentano. Viene spesso utilizzato per confrontare i valori di **diverse categorie di dati**.

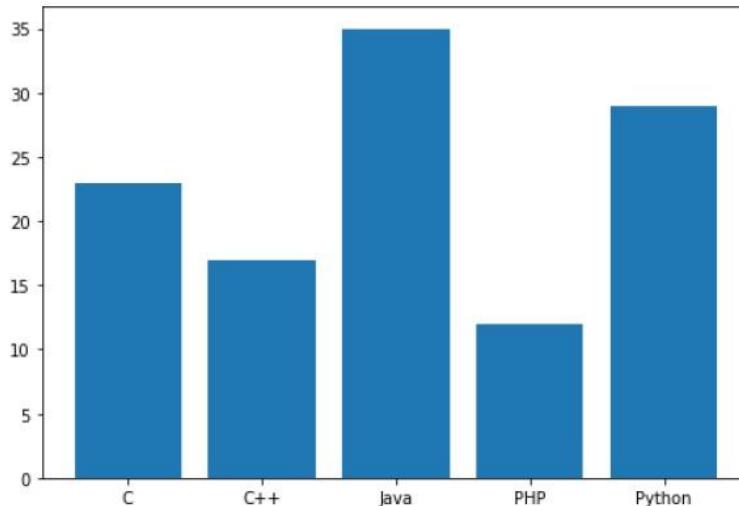
Un dato categoriale non è altro che un raggruppamento di dati in diversi gruppi logici, ad esempio, i dati sull'altezza delle persone raggruppate come "Alto", "Medio", "Basso" ecc. Per creare un grafico a barre, è necessario calcolare il totale di ciascuna categoria.

Un grafico a barre rappresenta una stima della **tendenza centrale** per una variabile

- numerica con l'altezza di ciascun rettangolo e fornisce alcune indicazioni dell'incertezza attorno a tale stima utilizzando le barre di errore. I grafici a barre includono lo 0 nell'intervallo dell'asse quantitativo e sono una buona scelta quando 0 è un valore significativo per la variabile quantitativa e si desidera fare confronti con esso.

# BARPLOT O BARCHART

È anche importante tenere presente che un grafico a barre mostra solo il **valore medio** (o altro stimatore), ma in molti casi può essere più informativo mostrare la distribuzione dei valori a ciascun livello delle variabili categoriali.



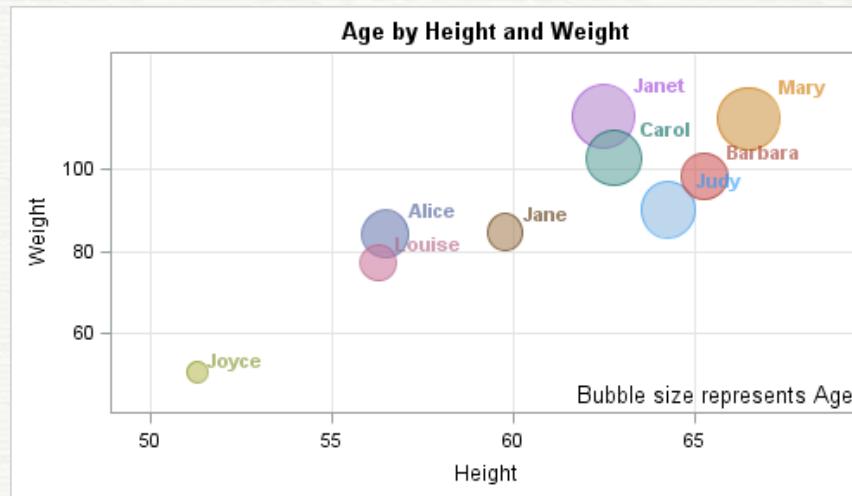
# BUBBLE CHART

Un diagramma a bolle (detto anche bubble chart o bostoniana) è un tipo di grafico in cui ogni entità rappresentata è definita in termini di **tre parametri numerici distinti**. I diagrammi a bolle sono usati per descrivere relazioni in vari ambiti scientifici.

Le entità rappresentate possono essere confrontate tra loro in base alla loro **dimensione** e alla loro **posizione** rispetto agli assi numerici. Infatti gli assi X e Y di un bubble chart sono scale numeriche, quindi la posizione in cui viene rappresentato un dato, descrive due valori numerici, mentre l'area del disegno dipende dal valore del terzo parametro.

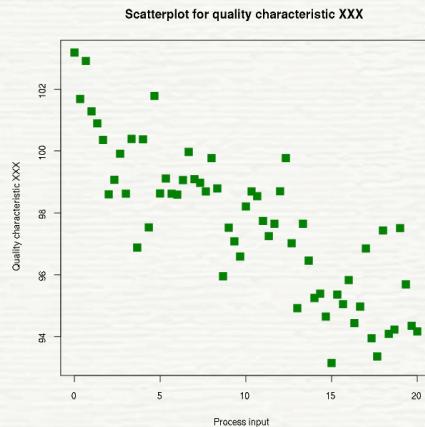
# BUBBLE CHART

Un bubble chart può essere considerato come una variazione di un grafico di dispersione dove i punti sono sostituiti con bolle. Questo tipo di grafico può essere utilizzato al posto di un grafico di dispersione se i dati presentano tre serie, ognuna delle quali contiene un insieme di dati.



# SCATTERPLOT

Per rappresentare una o più **variabili cardinali**, viene comunemente utilizzato il grafico a dispersione o scatterplot. Lo scatteplot è un grafico classico utilizzato per studiare la relazione tra **due variabili** riportate su un piano cartesiano. I dati sono visualizzati tramite una collezione di punti ciascuno con una posizione sull'asse orizzontale determinato da una variabile e sull'asse verticale determinato dall'altra. Se si hanno gruppi multipli di dati nel dataset, è possibile visualizzare ogni gruppo in un colore diverso.



# PARALLEL CATEGORIES DIAGRAM

Il Parallel categories diagram (noto anche come set parallelo) è una visualizzazione di set di **dati categorici multidimensionali**. Ogni variabile nel set di dati è rappresentata da una colonna di rettangoli, dove ogni **rettangolo** corrisponde a un **valore discreto** assunto da quella variabile. Le **altezze** relative dei rettangoli riflettono la **frequenza relativa** di occorrenza del valore corrispondente.

Le combinazioni di rettangoli di categoria di più dimensioni sono collegate da **nastri**, dove l'altezza del nastro corrisponde alla frequenza relativa di occorrenza della combinazione di categorie nel set di dati.

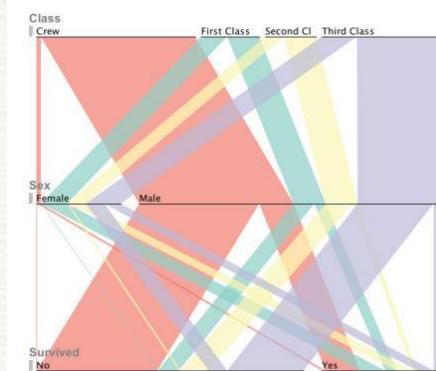
Questo tipo di visualizzazione mostra i possibili insiemi e sottoinsiemi esistenti nei dati, nonché le loro dimensioni.

# PARALLEL CATEGORIES DIAGRAM

La prima versione del Parallel Sets era basata sulle categorie e sugli incroci. Per ogni asse abbiamo mostrato ogni categoria come una casella e la sua dimensione era corrispondente alla frazione di tutti i punti dati rappresentati da ciascuna categoria.

In termini di statistica, questa è chiamata **distribuzione marginale** (o probabilità marginale).

Ogni asse è essenzialmente un grafico a barre, con le barre ribaltate. I nastri collegano le categorie di eventi che si verificano insieme, mostrando la frequenza delle classi.



# HERTZSPRUNG-RUSSELL DIAGRAM

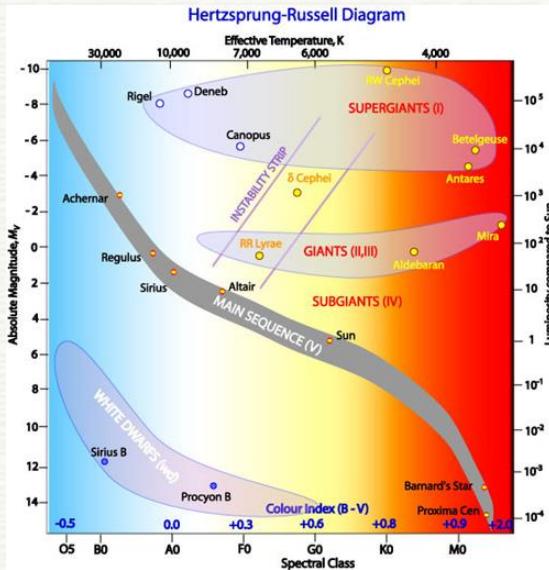
Il diagramma di Hertzsprung-Russel o diagramma HR è stato creato da all'inizio del 1900 ed è uno dei tool più importanti per lo studio dell'evoluzione stellare. È stato ideato dai due astronomi Ejnar Hertzsprung e Henry Norris Russell.

È un diagramma molto importante per lo studio delle stelle. Solitamente viene utilizzato per **analizzare il rapporto** che intercorre fra le seguenti coppie di valori: **absolute magnitude e colore** oppure **classi spettrali e absolute magnitude** oppure **temperatura della stella e luminosità**. Il diagramma ottenuto è in grado di mostrare che la relazione fra temperatura e luminosità non è casuale ma vengono a crearsi dei gruppi ben delimitati.

- Ciascuna stella, a seconda anche della propria massa e del modo in cui produce energia, viene assegnata ad uno dei gruppi che si sono così creati.

# HERTZSPRUNG-RUSSELL DIAGRAM

Per un astronomo è semplice ricavare informazioni da questo diagramma: è possibile ricavare la **struttura interna di una stella e lo stadio evolutivo** semplicemente determinandone la posizione nel diagramma. Questo perché all'interno del diagramma sono mostrate le stelle in diversi stadi della propria evoluzione.





# ANALISI DEL DATASET

---

[Link](#) al notebook  
[Link](#) al notebook per small multiple

# INFORMAZIONI SUL DATASET

All'interno del dataset sono contenute 7 colonne, ciascuna corrispondente ad un diverso attributo. Gli attributi sono rispettivamente:

1. **Temperatura del colore**, viene espressa in Kelvin.
2. **L**, è la luminosità della stella cioè la quantità di energia elettromagnetica che viene emessa dalla stella stessa per unità di tempo, ovvero la sua potenza. Viene espresso in relazione alla luminosità del sole.
3. **R**, è il raggio della stella. Viene espresso in relazione al raggio del sole.
4. **AM**, parametro che dipende dalla luminosità intrinseca della stella e dalla sua distanza con la terra.

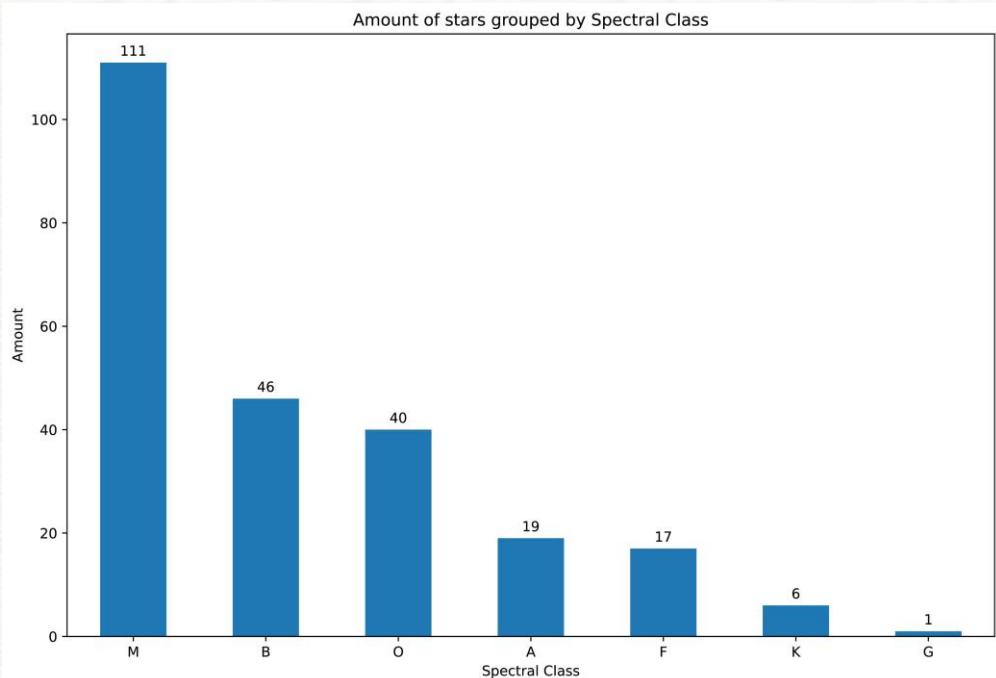
# INFORMAZIONI SUL DATASET

5. **Colore**, indica il colore dello spettro della stella.
6. **Classe spettrale** a cui appartengono le stelle.
7. **Tipo di stella**, viene assegnata considerando i parametri di colore, spettro di emissione ed albedo.

- Red Dwarf
- Brown Dwarf
- White Dwarf
- Main Sequence
- Super Giants
- Hyper Giants

	Temperature	L	R	A_M	Color	Spectral_Class	Type
0	3068	0.002400	0.1700	16.12	Red	M	Red Dwarf
1	3042	0.000500	0.1542	16.60	Red	M	Red Dwarf
2	2600	0.000300	0.1020	18.70	Red	M	Red Dwarf
3	2800	0.000200	0.1600	16.65	Red	M	Red Dwarf
4	1939	0.000138	0.1030	20.06	Red	M	Red Dwarf
...	...	...	...	...	...	...	...
235	38940	374830.000000	1356.0000	-9.93	Blue	O	Hyper Giants
236	30839	834042.000000	1194.0000	-10.63	Blue	O	Hyper Giants
237	8829	537493.000000	1423.0000	-10.73	White	A	Hyper Giants
238	9235	404940.000000	1112.0000	-11.23	White	A	Hyper Giants
239	37882	294903.000000	1783.0000	-7.80	Blue	O	Hyper Giants

# INFORMAZIONI GENERALI



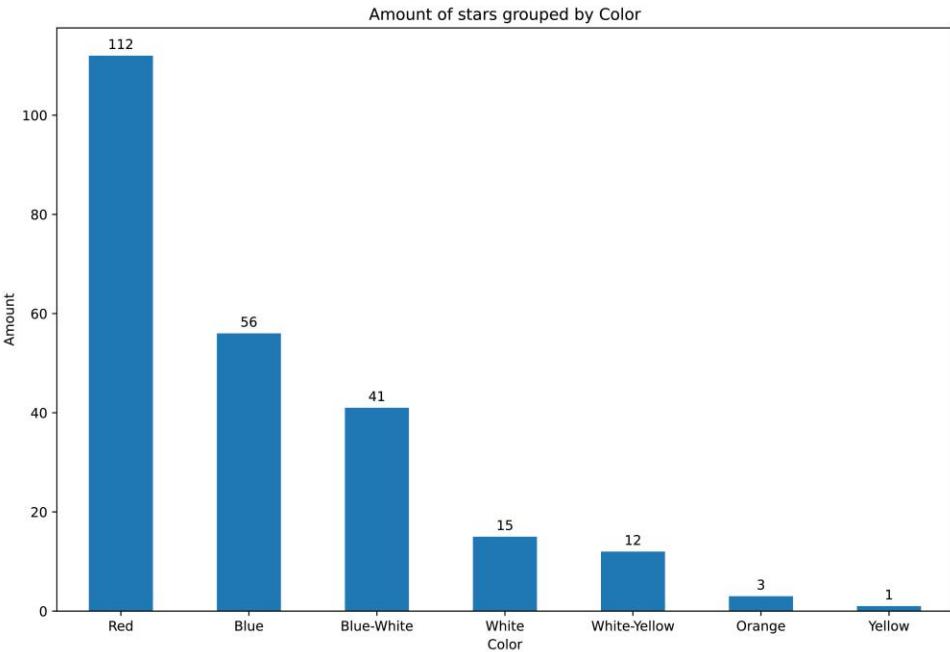
Tramite l'utilizzo di un bar chart vengono mostrate le proporzioni tra le sette tipologie di classi spettrali. La classe spettrale di una stella si identifica spesso con la sua temperatura di colore. Possiamo classificare le stelle appartenenti alla classe O come quelle più fredde, dall'altra parte invece la classe spettrale M rappresenta gli astri più caldi. Le classi si identificano con:

- O 28000 - 50000 K
- B 10000 - 28000 K
- A 7500 - 10000 K
- F 6000 - 7500 K
- G 4900 - 6000 K
- K 3500 - 4900 K
- M < 3500 K

Concludiamo che le stelle più calde e le stelle più fredde sono quelle prevalenti.

Le stelle nell'intorno più interno allo spettro sono proporzionalmente inferiori

# INFORMAZIONI GENERALI

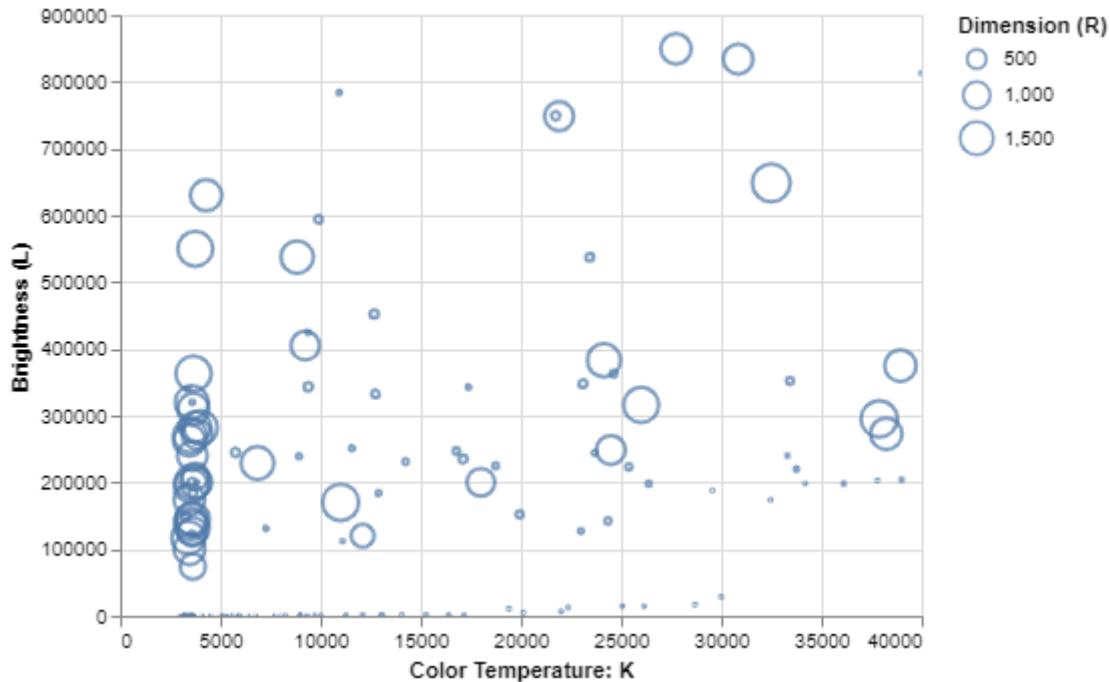


Analogamente al grafico precedente mostriamo la distribuzione delle stelle basandoci sul colore dell'astro celeste. Come vediamo possiamo suddividere i colori in sette categorie, ciascuno associato ad una classe spettrale:

- O: Blu
- B: Bianco-Blu
- A: Bianco
- F: Bianco-Giallo
- G: Giallo
- K: Arancione
- M: Rosso

Vi è dunque una stretta relazione tra colore e classe spettrale.

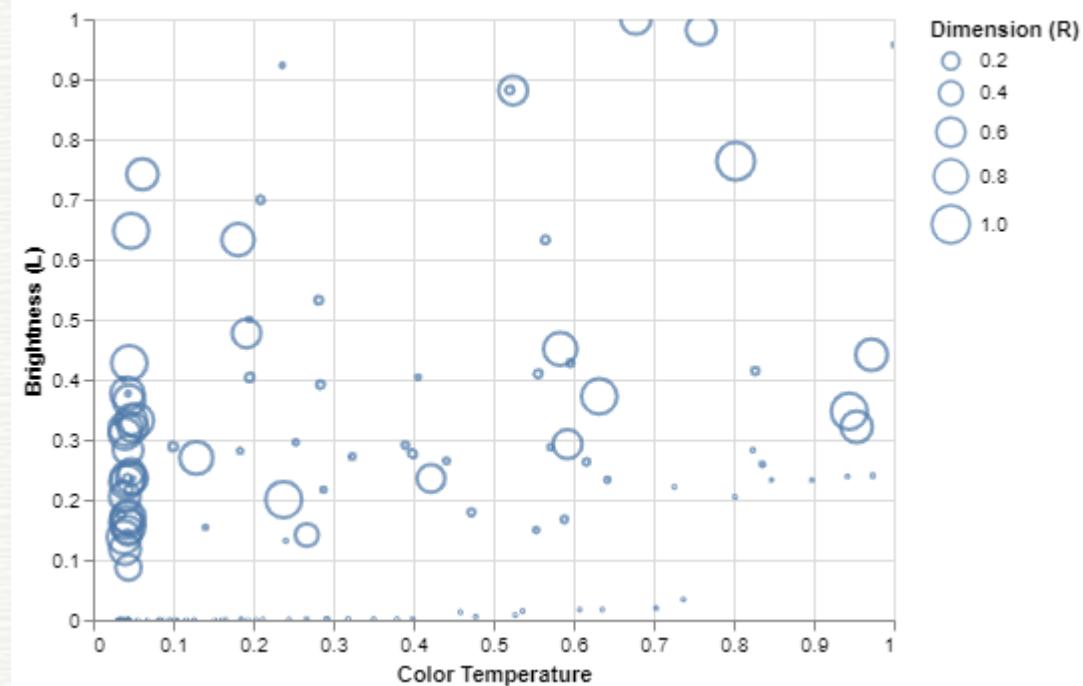
# RAPPRESENTAZIONE DEL DATASET



Dalla rappresentazione del dataset non possiamo ricavare molte informazioni ma possiamo vedere come sono distribuiti i dati.

La scelta di lasciare le griglie è dovuta all'interattività del grafico sul notebook

# IL GRAFICO NORMALIZZATO



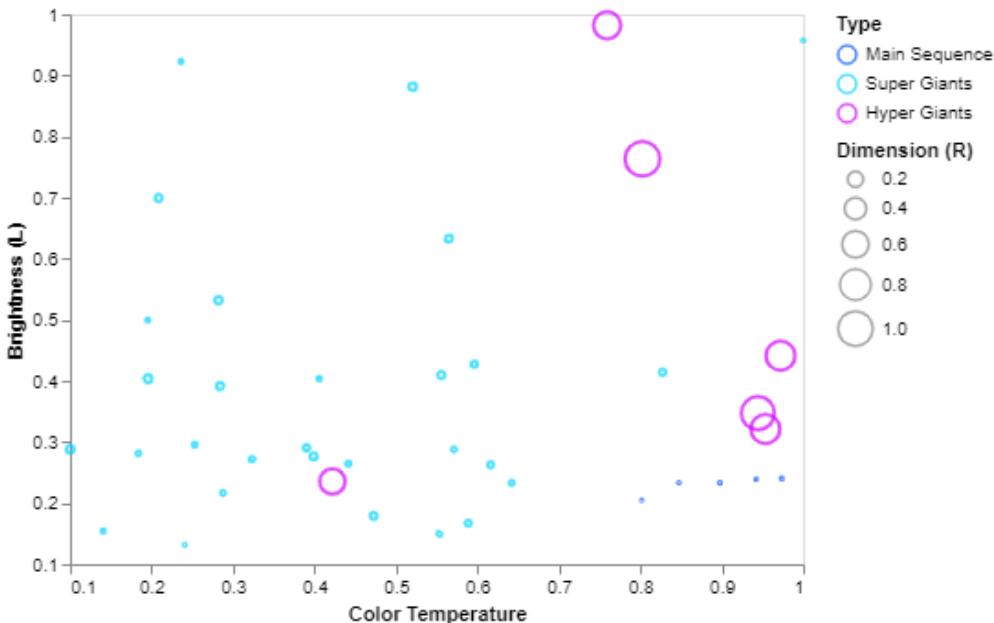
Come il grafico precedente, nemmeno creando il grafico coi dati normalizzati ci è possibile ricavare informazioni particolarmente rilevanti

# **ANALISI DELLA DISTRIBUZIONE PER CLASSE SPETTRALE**

Nei grafici presenti nelle slide successive sono stati creati dei bubble plot suddividendo le stelle per classi spettrali.

I colori usati per il tipo di stella e la legenda sulle dimensioni sono uguali per tutti i grafici così da poter seguire le regole per la buona rappresentazione dei dati

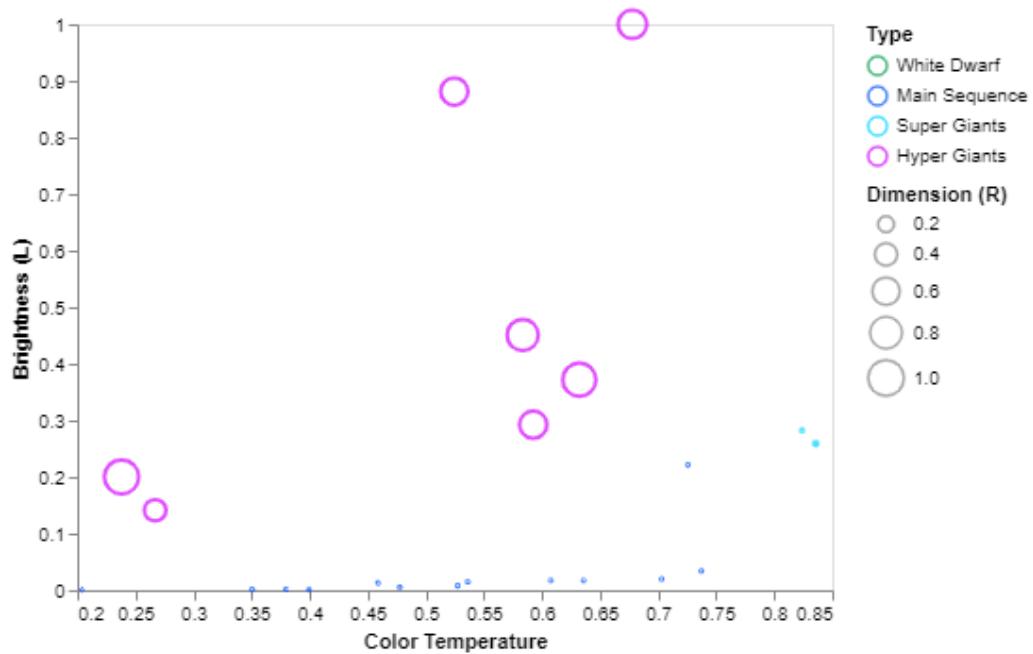
# DISTRIBUZIONE NELLA CLASSE O



In questo primo grafico possiamo vedere come in questa classe spettrale siano presenti tre tipi diversi di stelle di dimensione, luminosità e temperatura di colore variabile.

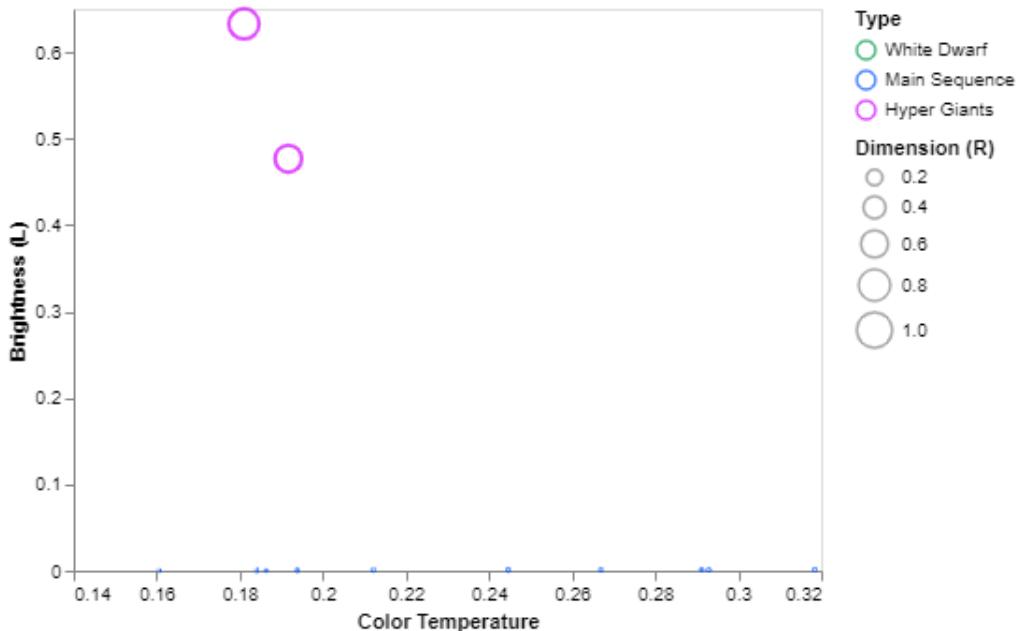
Possiamo notare come sia presente un cluster di stelle appartenenti al gruppo **Main Sequence** nella parte inferiore destra del grafico: sono stelle con una luminosità bassa ma una temperatura elevata. È possibile altresì notare che sono stelle molto piccole. Sempre nella parte più a destra del grafico, quindi fra le stelle con una temperatura elevata, è possibile trovare le stelle appartenenti al gruppo delle **Hyper Giants**. Possiamo notare come le stelle appartenenti a questo gruppo siano stelle di dimensioni elevate rispetto alle altre presenti in questa classe spettrale. Notiamo come le stelle appartenenti alla categoria delle **Super Giants** siano stelle piccole che hanno prevalentemente una temperatura e una luminosità medio-bassa.

# DISTRIBUZIONE NELLA CLASSE B



In questo grafico possiamo vedere come in questa classe spettrale siano presenti quattro tipi diversi di stelle di dimensione, luminosità e temperatura di colore variabile. Bisogna notare come la temperatura di tutte le stelle di questa classe spettrale non sia vicina allo zero e nemmeno vicina all'uno: si pongono nei valori centrali del dataset. Le stelle di dimensione maggiore appartengono al gruppo delle **Hyper Giants** mentre le stelle delle altre tre categorie sono molto piccole. A tal proposito è possibile notare come le **White Dwarf** siano talmente piccole da perdersi nel grafico, le **Super Giants** sono le stelle che si trovano più a destra del grafico e quindi le più calde mentre le **Main Sequence** hanno una luminosità bassissima, prossima allo zero.

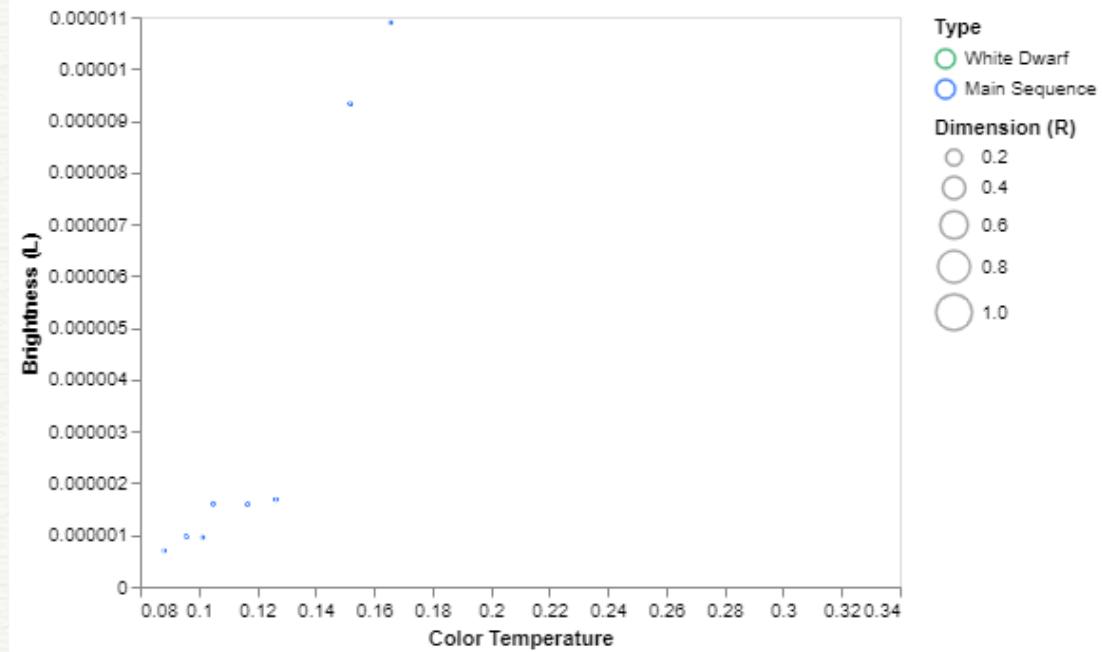
# DISTRIBUZIONE NELLA CLASSE A



In questo grafico possiamo vedere come in questa classe spettrale siano presenti tre tipi diversi di stelle di dimensione, luminosità e temperatura di colore variabile. Bisogna notare come la temperatura di tutte le stelle di questa classe spettrale sia particolarmente bassa.

Le stelle di dimensione maggiore appartengono al gruppo delle **Hyper Giants** mentre le stelle delle altre categorie sono molto piccole. A tal proposito è possibile notare come le **White Dwarf** siano talmente piccole da perdersi nel grafico e le **Main Sequence** hanno una luminosità bassissima, prossima allo zero.

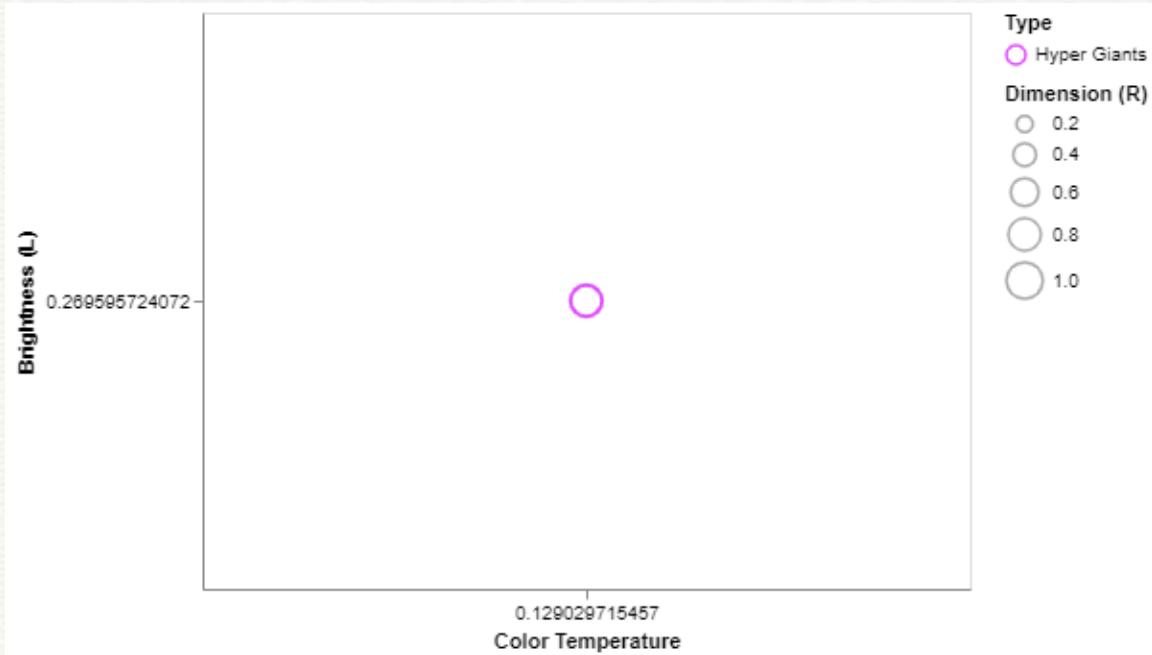
# DISTRIBUZIONE NELLA CLASSE F



In questo grafico possiamo vedere come in questa classe spettrale siano presenti due tipi diversi di stelle tutte di una dimensione molto piccola.

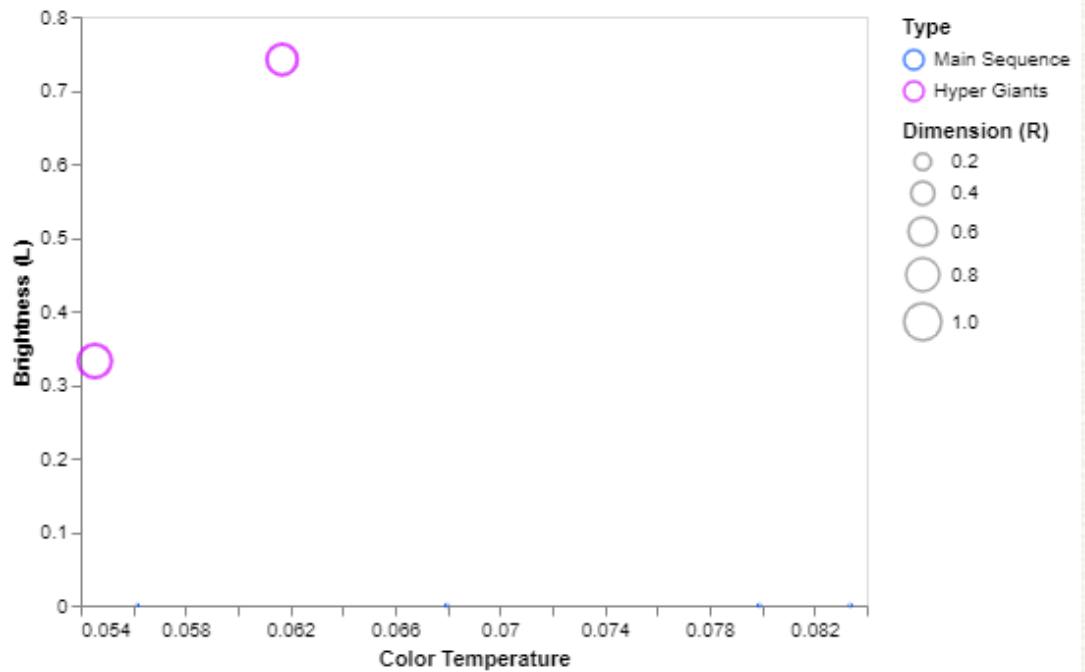
Le stelle di dimensione maggiore appartengono al gruppo delle **Main Sequence** e vediamo che, pur non essendo particolarmente luminose, sono più luminose delle **White Dwarf**, la cui luminosità è possima allo zero

# DISTRIBUZIONE NELLA CLASSE G



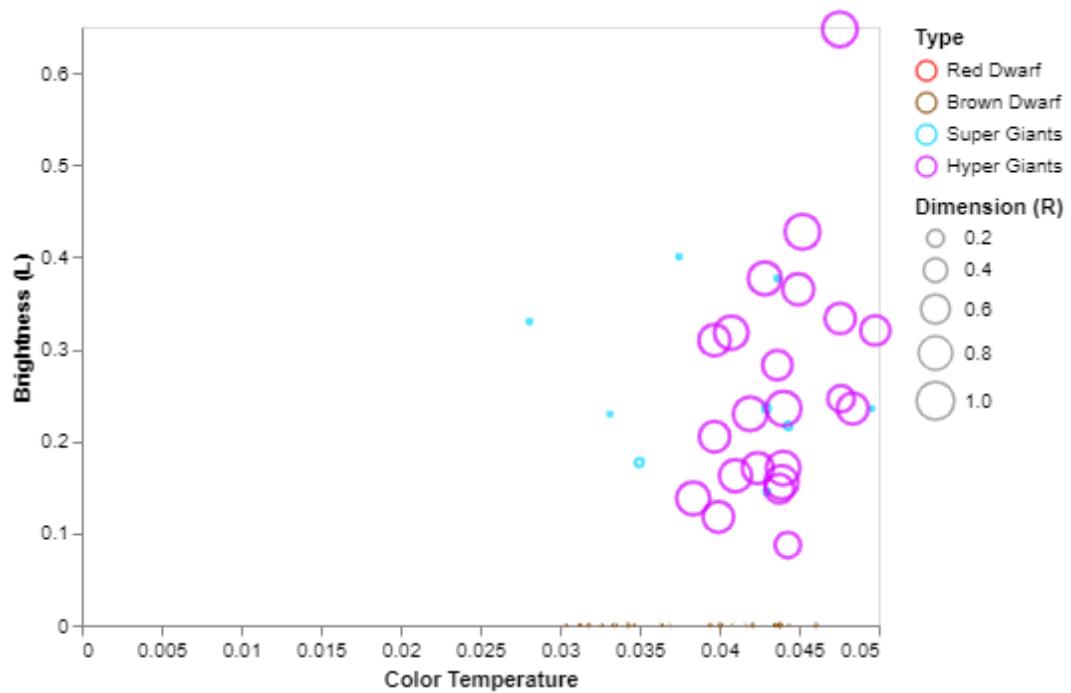
A questa classe spettrale appartiene una sola stella: non è possibile ricavare dati esaustivi

# DISTRIBUZIONE NELLA CLASSE K



In questo grafico possiamo vedere come in questa classe spettrale siano presenti due tipi diversi di stella.  
Le stelle del tipo **Hyper Giants** sono più grandi e più luminose delle **Main Sequence**

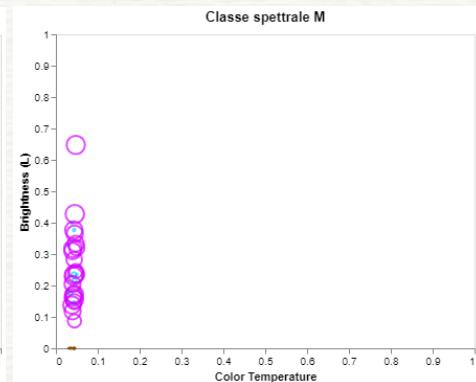
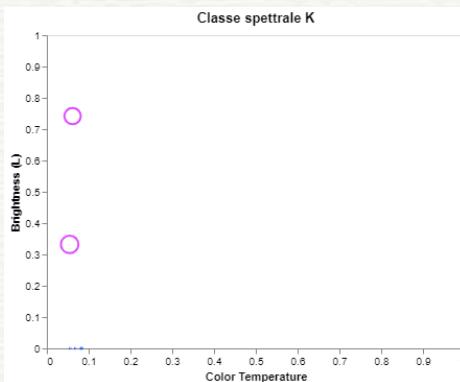
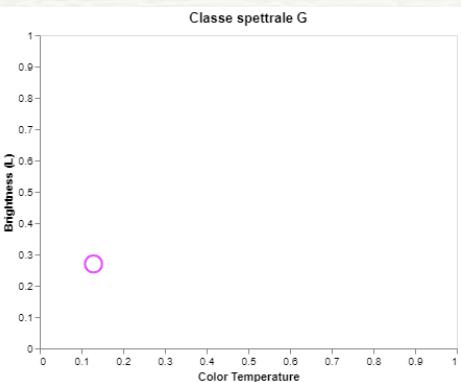
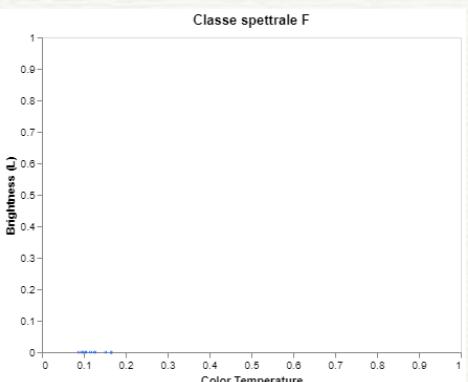
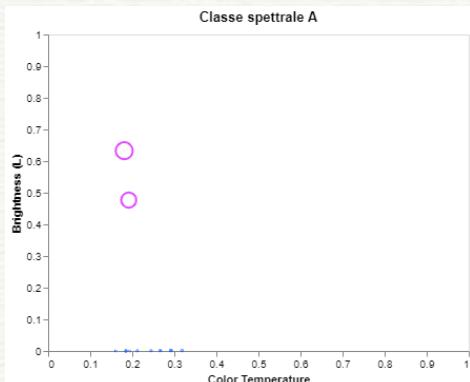
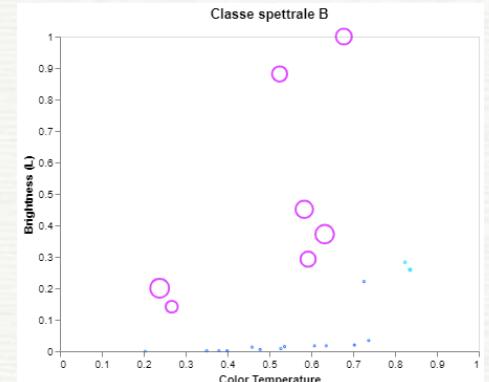
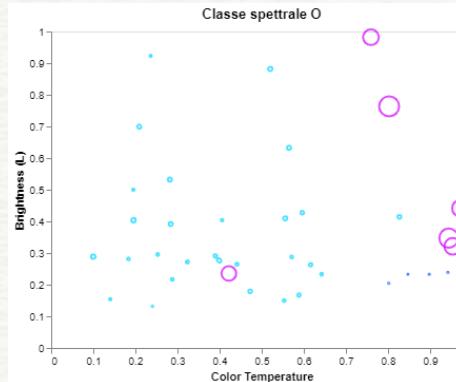
# DISTRIBUZIONE NELLA CLASSE M



In questo grafico possiamo vedere come in questa classe spettrale siano presenti quattro tipi diversi di stelle, differenti per dimensione, luminosità e temperatura di colore.

Notiamo che le stelle di dimensione maggiore sono quelle del tipo **Hyper Giants** che, pur avendo una temperatura bassa rispetto alla totalità delle stelle, sono quelle con colore più freddo della classe spettrale. Notiamo che sovrapposte per temperatura e luminosità alle Hyper Giants abbiamo le stelle appartenenti al tipo **Super Giants**. Con una luminosità prossima allo zero abbiamo le stelle appartenenti alle classi **Red Dwarf** e **Brown Dwarf** che, vista la loro dimensione davvero piccola, quasi si perdono all'interno del grafico.

# SMALL MULTIPLE: IL GRAFICO



# SMALL MULTIPLE: COMMENTI

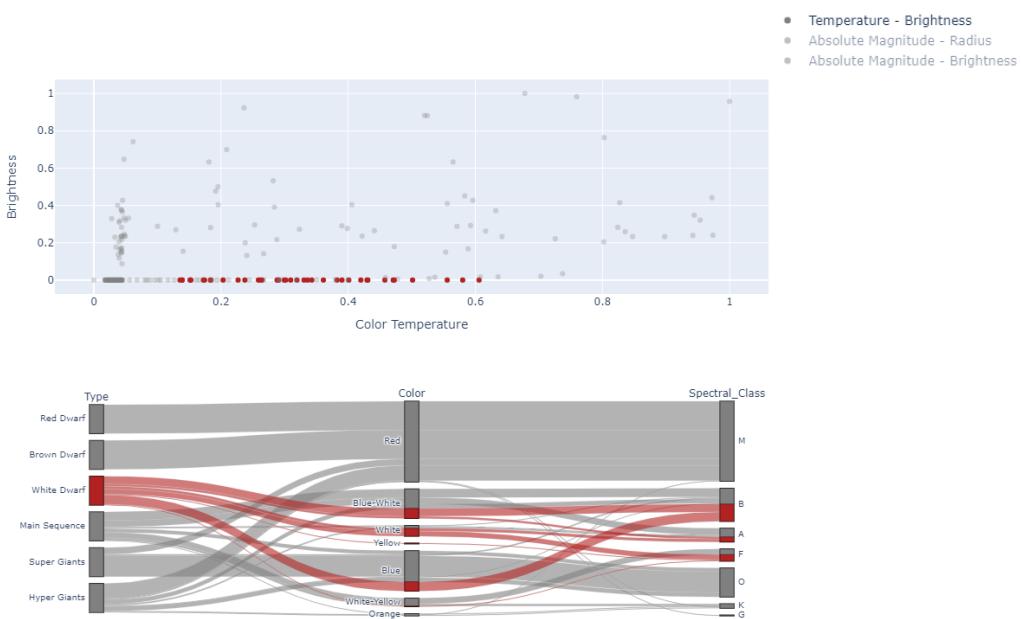
Nella slide precedente è stato creato uno small multiple per poter comparare la distribuzione delle stelle all'interno delle singole classi spettrali. A differenza delle rappresentazioni fatte sulla singola classe spettrale, dove si è deciso di zoommare solo sull'area dove sono effettivamente presenti i dati, qui sono state mantenute le dimensioni complete del dataset. Anche in questo caso sono stati utilizzati i dati normalizzati per creare i grafici. I grafici sono stati rappresentati in ordine di classe dalla più fredda alla più calda: O, B, A, F, G, K, M

La legenda sui colori e la dimensione delle bolle, essendo uguale per tutti i grafici, è stata mantenuta solamente nel primo.

La cosa che è possibile notare dalla comparazione di grafici è:

- Non in tutti i grafici sono presenti gli stessi tipi di stelle
- Avendo mantenuto la stessa scala per tutti i grafici, dove sono presenti stelle molto piccole queste si vanno a perdere
- In alcune classi le stelle si distribuiscono in un punto preciso del dataset mentre in altre sono presenti in tutti i punti

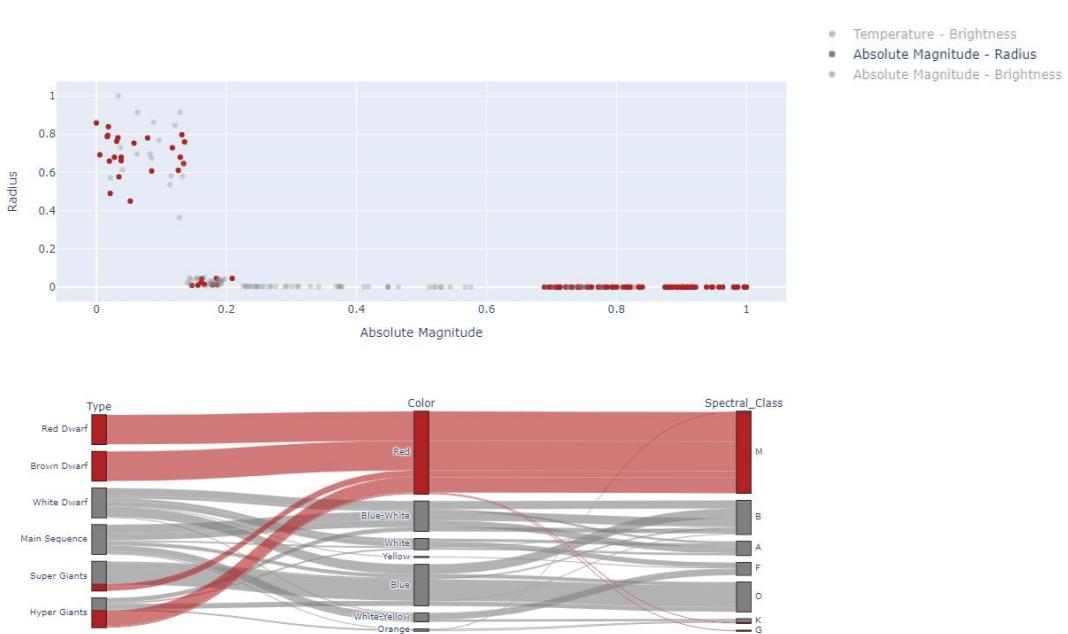
# SCATTERPLOT E PARALLEL DIAGRAM



Questa metodologia di rappresentazione di dati, chiamata **Parallel Categories Diagram**, è molto interessante e funzionale per mettere in relazione diversi dati. Il principale scopo di questo grafico è quello di analizzare la distribuzione dei dati nel dataset di riferimento.

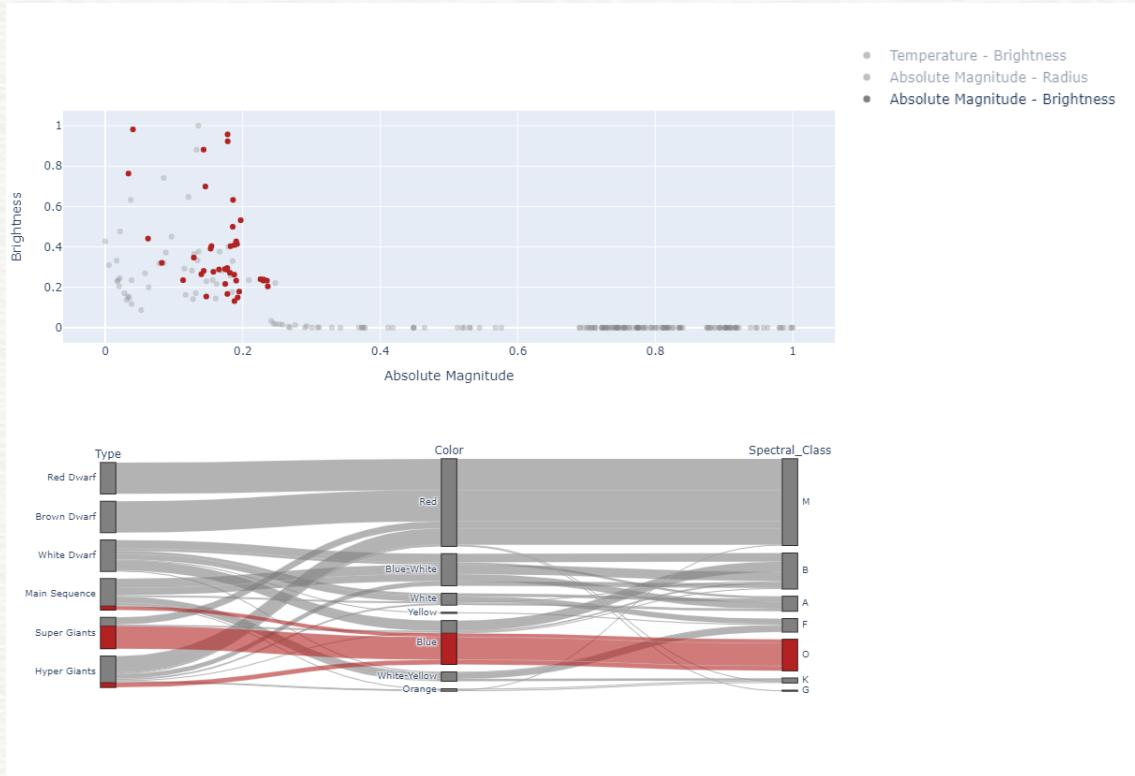
Questo rappresentazione tuttavia accetta solamente dati di tipo categorico e non numerico, infatti nel nostro caso abbiamo focalizzato la ricerca su tre dati quali il **Tipo** di stella, il suo **Colore** e la sua **Classe Spettrale**. Da qui possiamo approfondire le relazioni che abbiamo constatato precedentemente relative alla connessione tra classe spettrale e colore.

# SCATTERPLOT E PARALLEL DIAGRAM

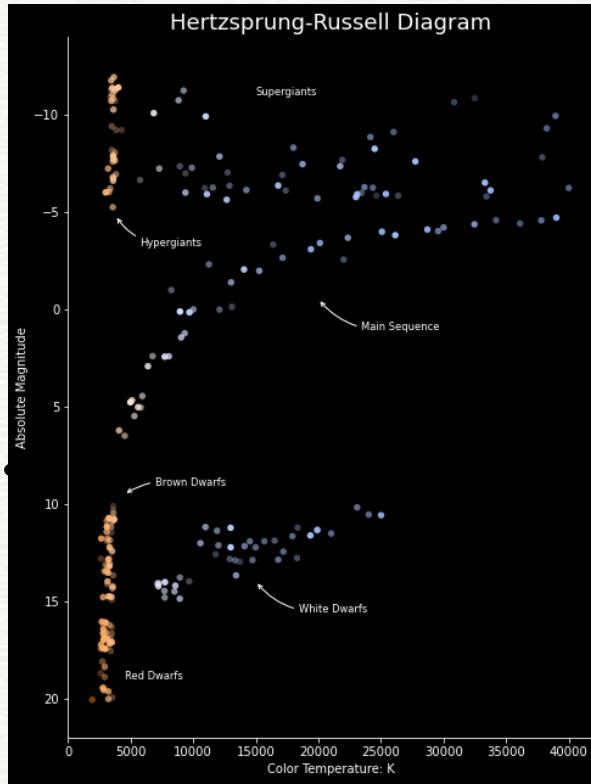


Abbiamo sviluppato tre Scatterplot diversi mettendo in relazione per ognuno dei dati differenti: il primo riguarda **temperatura e luminosità**, il secondo **magnitudine assoluta e raggio** e infine **magnitudine assoluta e luminosità**. Tutti e tre fanno riferimento allo stesso Parallel Categories Diagram. Questo ci è stato utile per approfondire le nostre conoscenze sulle varie caratteristiche che ogni categoria di stelle possiede, andando ad agire su tipologia di stella, sul suo colore o sulla sua classe spettrale.

# SCATTERPLOT E PARALLEL DIAGRAM



# HERTZSPRUNG-RUSSELL DIAGRAM



Questo grafico è chiamato **Hertzsprung-Russell Diagram** ed è usato molto spesso in astronomia per la classificazione degli astri celesti.

È un grafico molto comodo perché gli astronomi grazie a questo grafico su 4 assi riescono a classificare in modo esaustivo le varie tipologie di stelle.

Nel nostro caso il grafico contiene due assi quali la magnitudine assoluta e la temperatura di colore. La parte interessante del grafico è il colore. Infatti siamo andati a mappare la temperatura della stella in una codifica `rgba` che è il colore che viene poi visualizzato nel grafico per ogni stella.

Siamo infine andati ad aggiungere delle etichette sul grafico in modo tale da isolare le varie tipologie di stelle.

# SITOGRAFIA

- Dataset: <https://www.kaggle.com/brsdincer/star-type-classification>
- Python libraries: <https://matplotlib.org/> <https://altair-viz.github.io/>  
<https://ipywidgets.readthedocs.io/en/stable/> <https://plotly.com/>  
<https://pandas.pydata.org/>
- HR diagram: <https://astronomy.swin.edu.au/cosmos/h/hertzsprung-russell+diagram>  
[https://www.atnf.csiro.au/outreach/education/senior/astrophysics/stellarevolution\\_hrintro.html](https://www.atnf.csiro.au/outreach/education/senior/astrophysics/stellarevolution_hrintro.html)
- Principi della Gestalt: <https://www.toptal.com/designers/ui/gestalt-principles-of-design#:~:text=There%20are%20six%20individual%20principles,gestalt%2C%20such%20as%20common%20fate.>
- Tipi di dato: lezioni di statistica e analisi dei dati
- Normalizzazione dei dati: <https://www.lorenzogovoni.com/ridimensionamento-dei-dati>

# SITOGRAFIA

- Scatterplot: <https://www.agnesevardanega.eu/wiki/r/grafici/scatterplot>
- Barplot: <https://seaborn.pydata.org/generated/seaborn.barplot.html>
- Bubble chart: <https://chartio.com/learn/charts/bubble-chart-complete-guide/>
- Altri grafici: <https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-masterplots-python/>
- L'uso del colore: <https://data-storytelling.it/visualizzazione-efficace/gli-attributi-preattentivi-visualizzazione-efficace-colore/>
- Informazioni sulle stelle: <https://www.astronomia.com/risorse/approfondimenti/classifichiamo-le-stelle/classi-spettrali-e-classi-di-luminosita/>
- Ink ratio: <https://speedwell.com.au/en/insights/2019/the-manifesto-of-the-data-ink-ratio#:~:text=The%20data%2Dink%20ratio%20is,non%2DData%2DInk>.
- Altre informazioni non specificate in precedenza: slide delle lezioni di visualizzazione scientifica