Star type classification

Case study (part 1)

Students:

Magda Kobusińska,

Szymon Skwarek,

Antoni Klorek

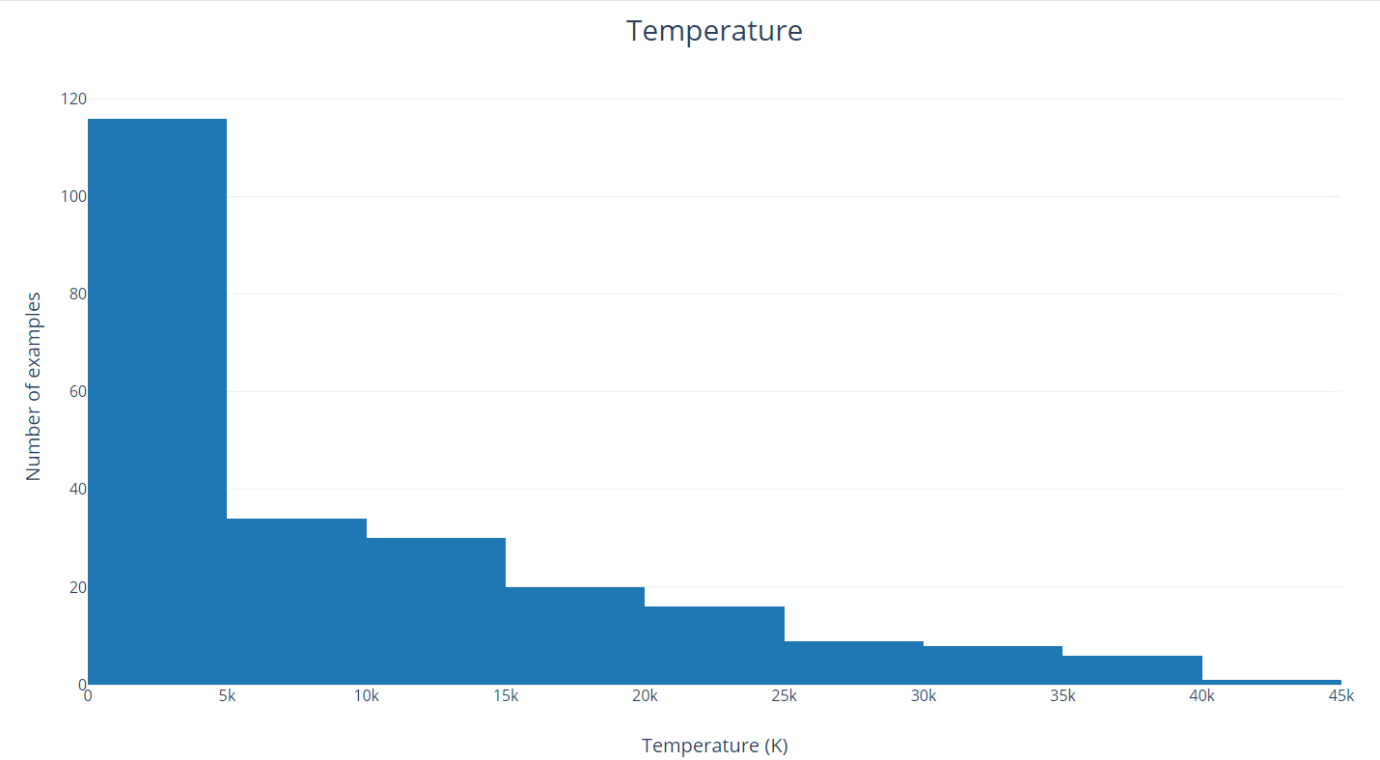
Dataset

- **Star type classification** dataset (data from NASA)
- **7 columns** (6 attributes and a target)
- **240 rows** (240 learning examples)
- **6** different **classification targets**

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
3692	0.00367	0.47	10.8	Red	M	1
3192	0.00362	0.1967	13.53	Red	M	1
3441	0.039	0.351	11.18	Red	M	1
25000	0.056	0.0084	10.58	Blue White	B	2
7740	0.00049	0.01234	14.02	White	A	2
7220	0.00017	0.011	14.23	White	F	2
8500	0.0005	0.01	14.5	White	A	2
16500	0.013	0.014	11.89	Blue White	B	2

Head of the dataset

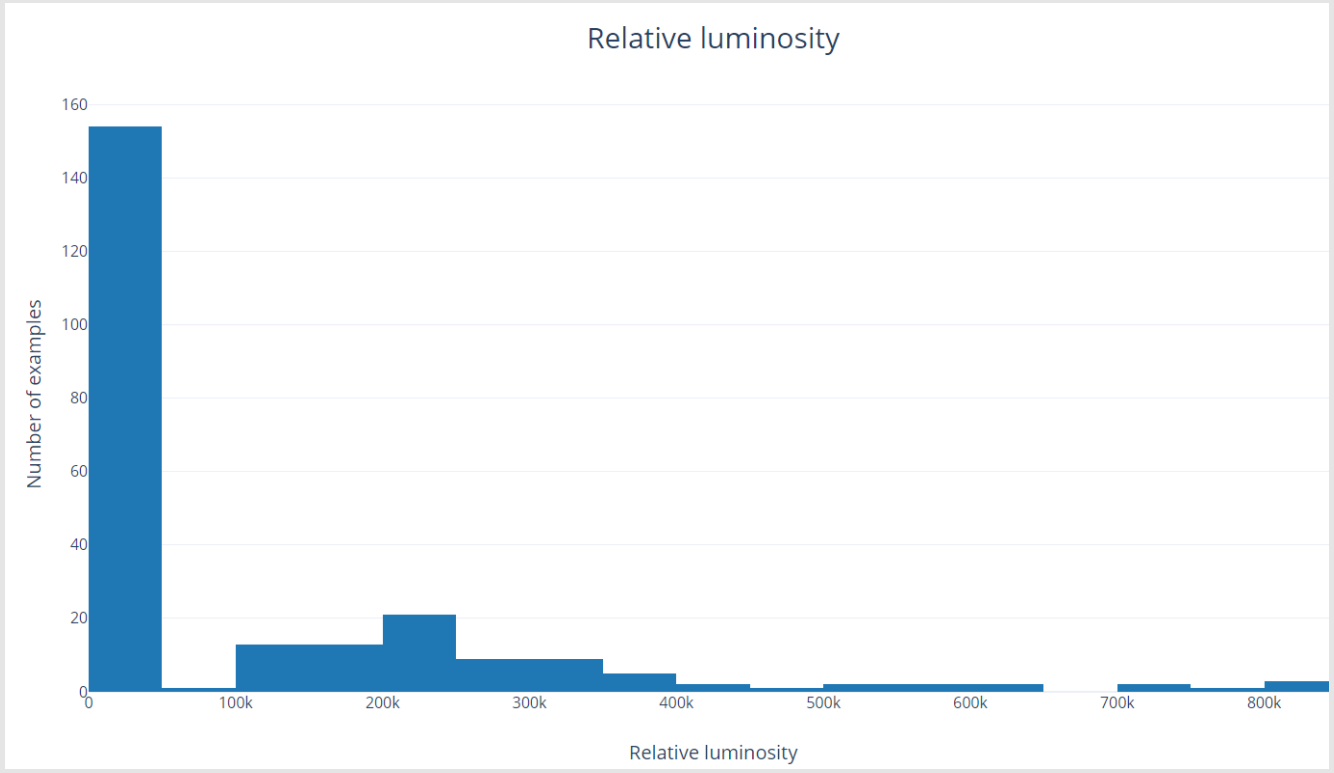
Attribute I: Temperature



- **Temperature** of a star in Kelvins
- Numeric type

- min = 1939K,
  max = 40000K
- mean = 10500K

Attribute II: L (Relative Luminosity)

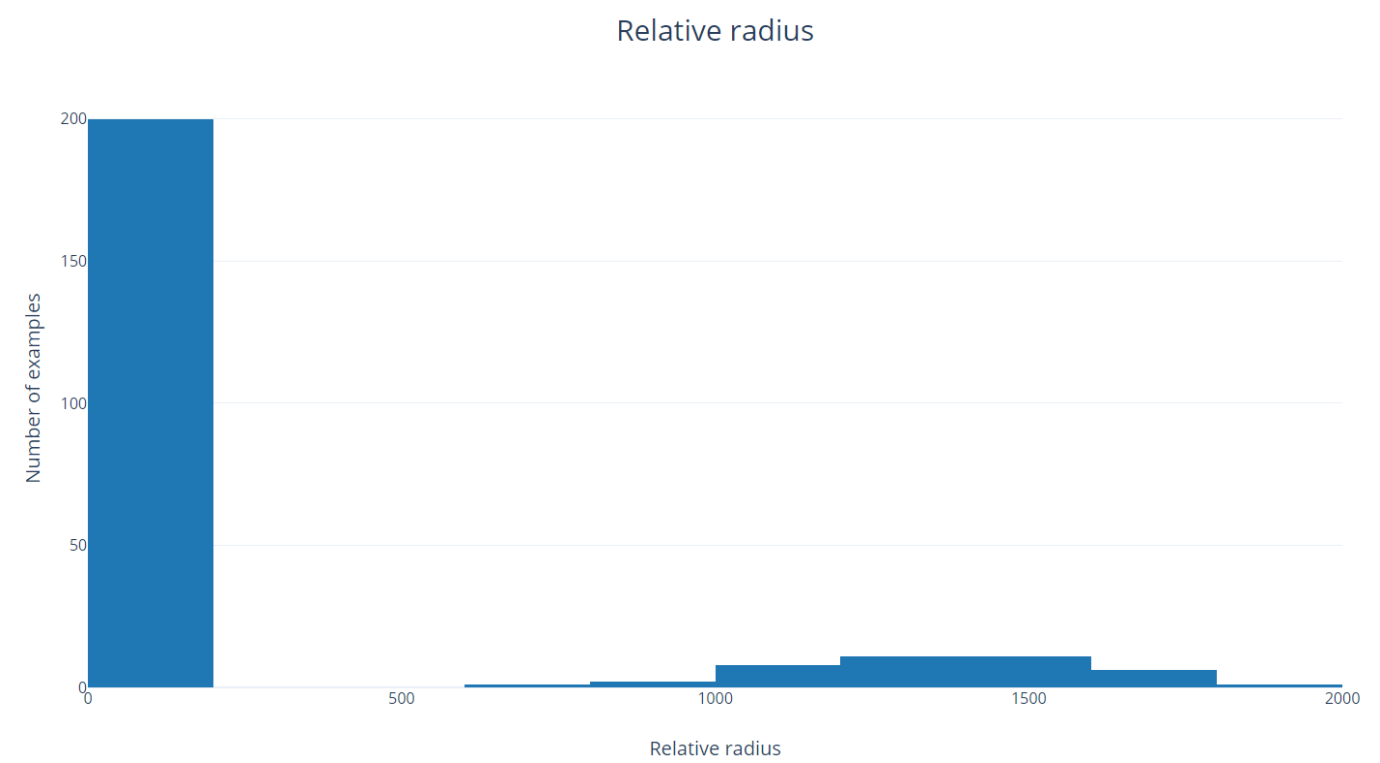


- **Luminosity** of a star devided by avg luminosity of the Sun (3.828*10^26 Watts)

Luminosity
the amount of energy (light) that a star emits from its surface

- **Numeric** type
- min = 0.00008,
  max = 849420
- mean = 107000

Attribute III: R (Relative Radius)

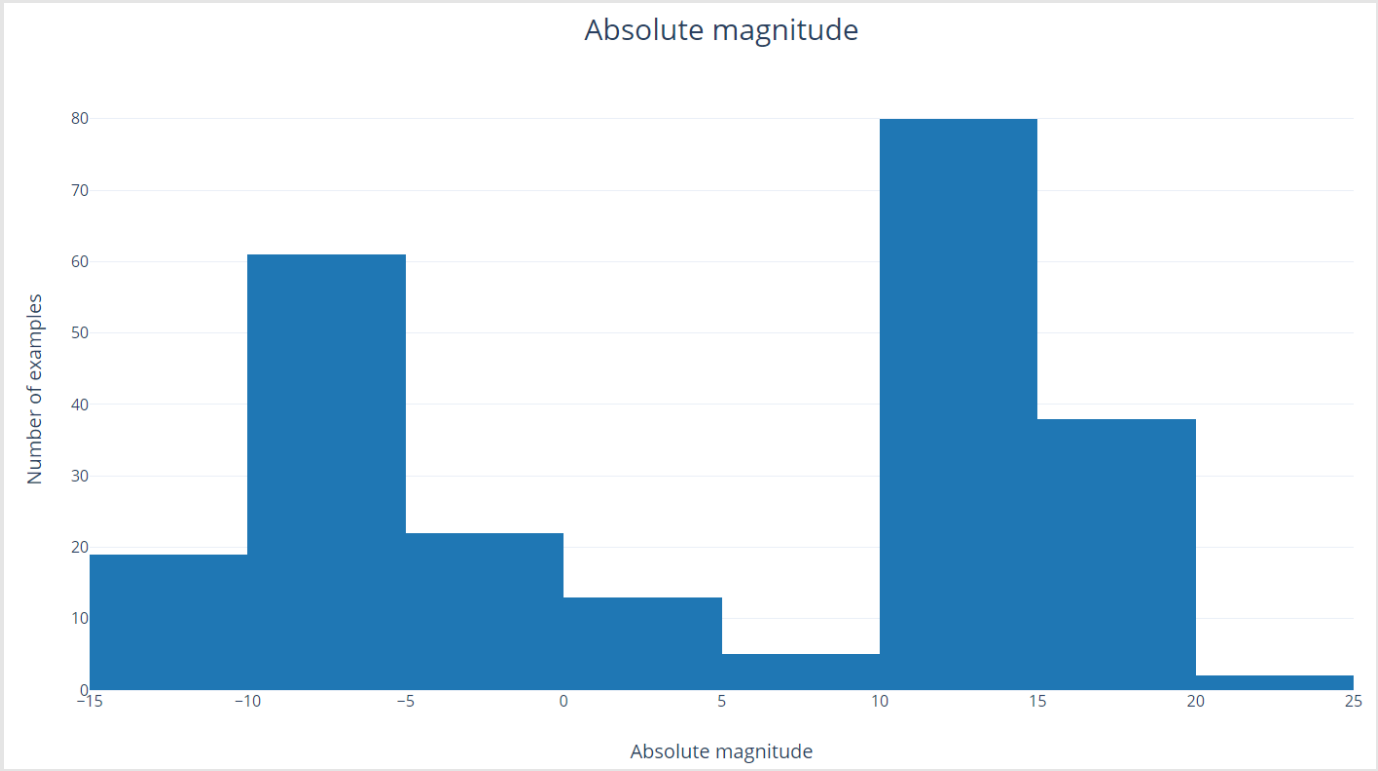


- **Radius** of a star devided by avg radius of the Sun (6.9551*10^8 m)
- **Numeric** type

- min = 0.0084,

  max = 1948.5
- mean = 237

Attribute IV: A_M (Absolute Magnitude)

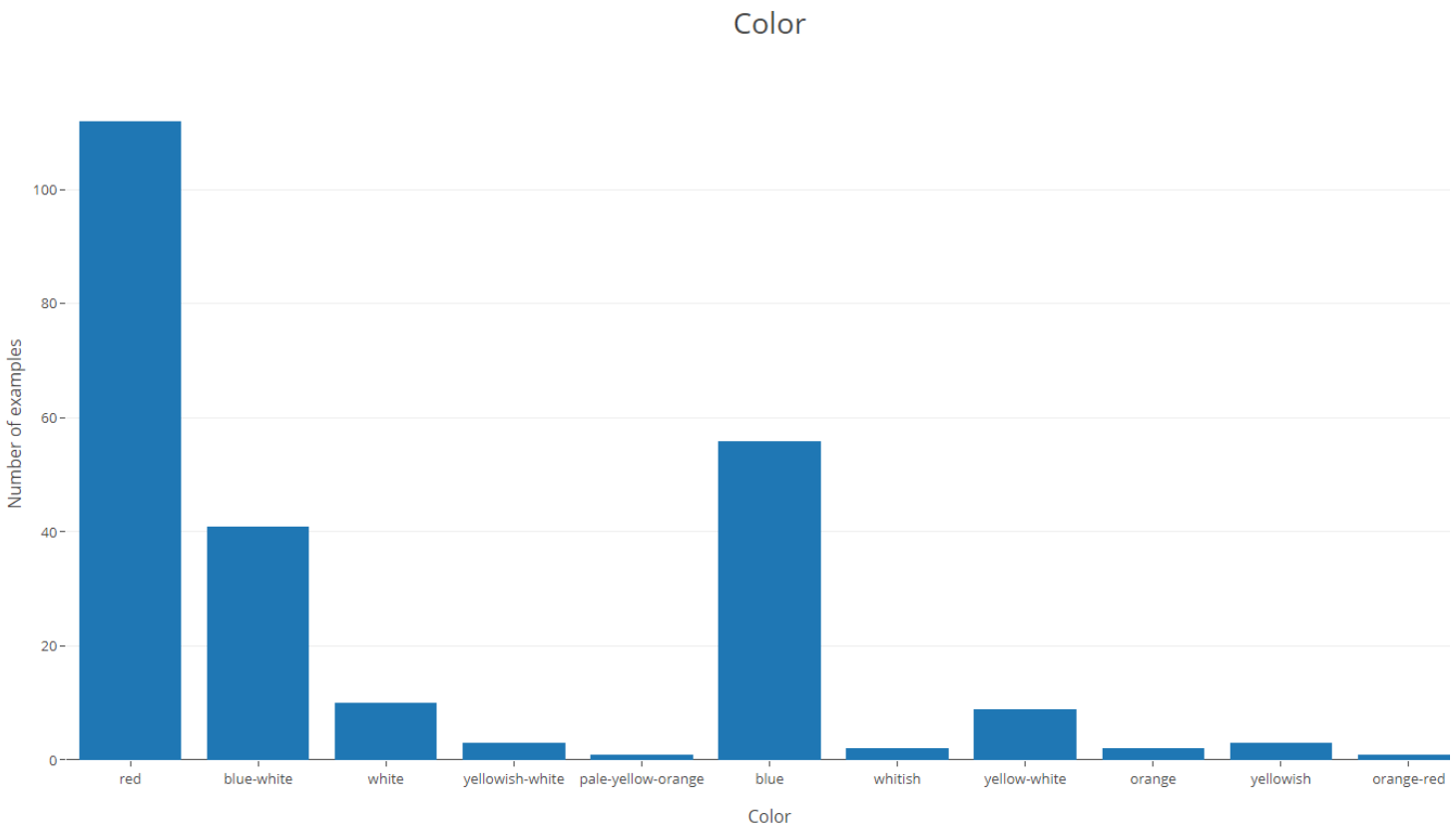


- **Absolute magnitude** how bright the star appears at a standard distance of 32.6 light years.

The more luminous an object, **the smaller** the numerical value of its absolute magnitude.

- **Numeric** type
- min = -11.92,
  max = 20.06
- mean = 4.38

Attribute V: Color



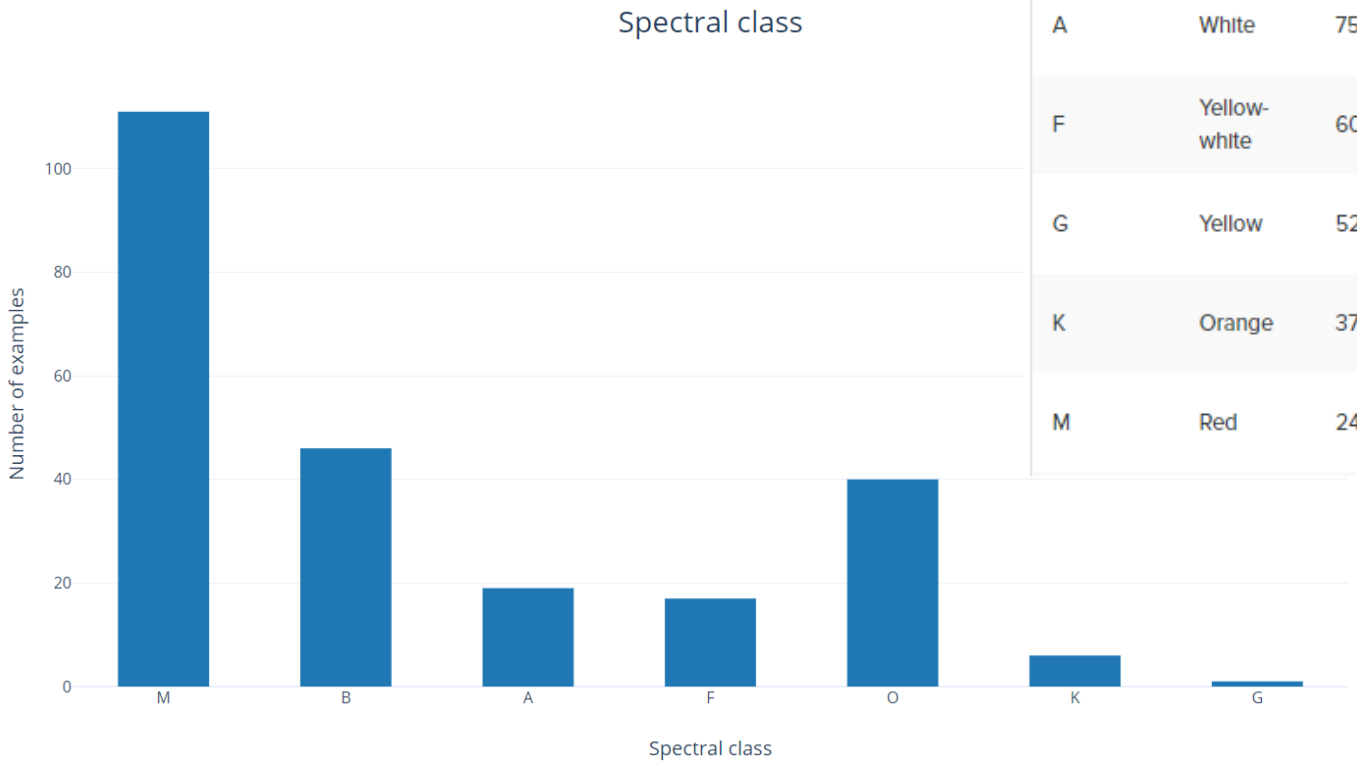
- General **Color of the Spectrum**
- Set of colors after unification:

  blue,
  blue-white,
  orange,
  orange-red,
  red,
  pale-yellow-orange,
  white,
  whitish,
  yellow-white,
  yellowish,
  yellowish-white

- **String** type
- Most common: **red (47%)**

Attribute VI: Spectral_Class

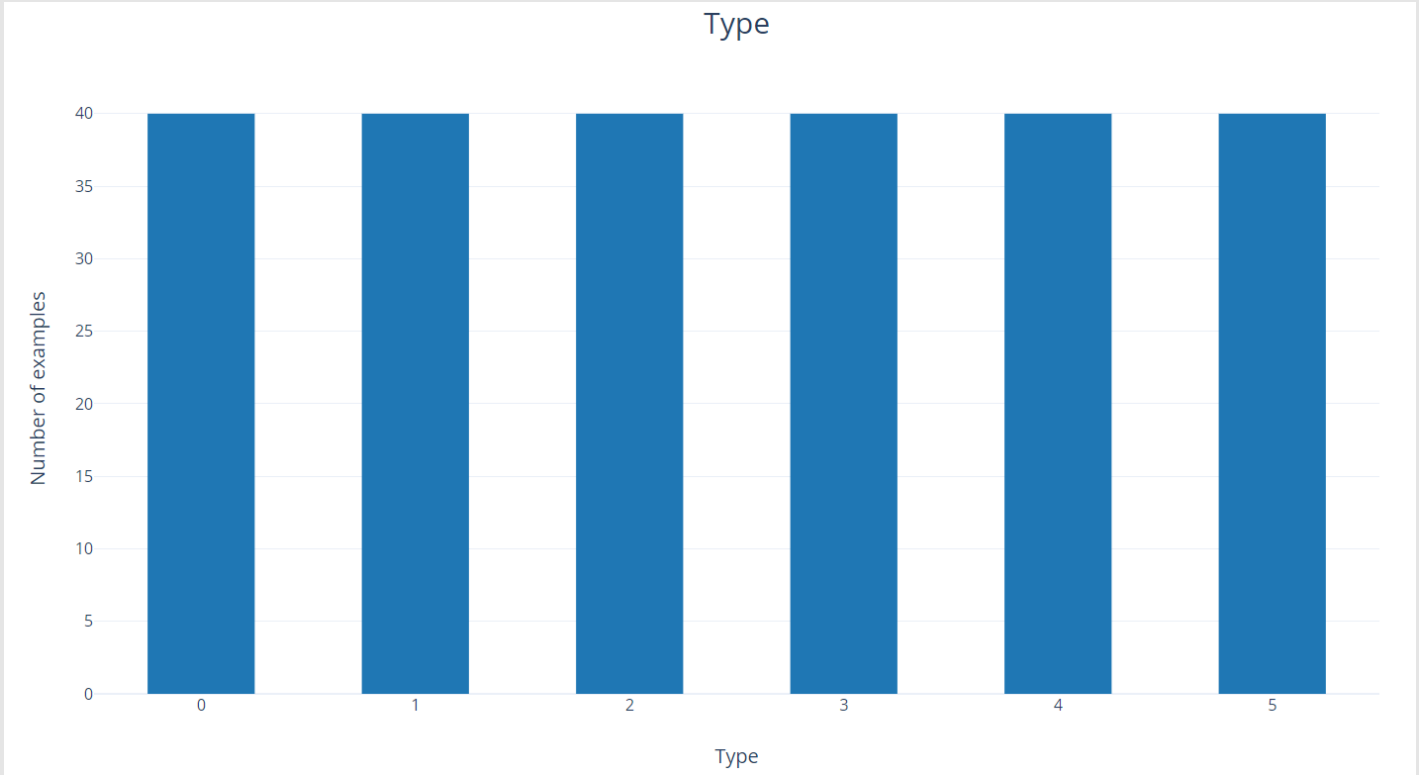
Set of **spectral class**:
{O, B, A, F, G, K, M}



Spectral Class	Color	Approximate Temperature (K)	Principal Features
O	Blue	> 30,000	Neutral and ionized helium lines, weak hydrogen lines
B	Blue-white	10,000–30,000	Neutral helium lines, strong hydrogen lines
A	White	7500–10,000	Strongest hydrogen lines, weak ionized calcium lines, weak ionized metal (e.g., iron, magnesium) lines
F	Yellow-white	6000–7500	Strong hydrogen lines, strong ionized calcium lines, weak sodium lines, many ionized metal lines
G	Yellow	5200–6000	Weaker hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of ionized and neutral metals
K	Orange	3700–5200	Very weak hydrogen lines, strong ionized calcium lines, strong sodium lines, many lines of neutral metals
M	Red	2400–3700	Strong lines of neutral metals and molecular bands of titanium oxide dominate

- **Character** type
- Most common: **M (46%)**

Target: Type



A **digit** from **0 to 5**:

- 0 - Red Dwarf
- 1 - Brown Dwarf
- 2 - White Dwarf
- 3 - Main Sequence
- 4 - Super Giants
- 5 - Hyper Giants

Correlations between numeric attributes

	Temperature	L	R	A_M	Type
Temperature	1.000000	0.393404	0.064216	-0.420261	0.411129
L	0.393404	1.000000	0.526516	-0.692619	0.676845
R	0.064216	0.526516	1.000000	-0.608728	0.660975
A_M	-0.420261	-0.692619	-0.608728	1.000000	-0.955276
Type	0.411129	0.676845	0.660975	-0.955276	1.000000

Data **cleaning**
&
preprocessing

Color naming **unification**

- Blue → blue
- Red → red
- White → white
- Orange → orange
- Whitish → whitish
- Yellowish → yellowish
- Orange-Red → orange-red
- White-Yellow → yellow-white
- Yellowish White → yellowish-white
- Pale Yellow Orange → pale-yellow-orange

- Blue white
- Blue White
- blue-White

} → blue-white