

**MINICURSO**  
**BÁSICO DE**  
  
**PETCC**

**R**

# Aula 03

Análise e Visualização de Dados

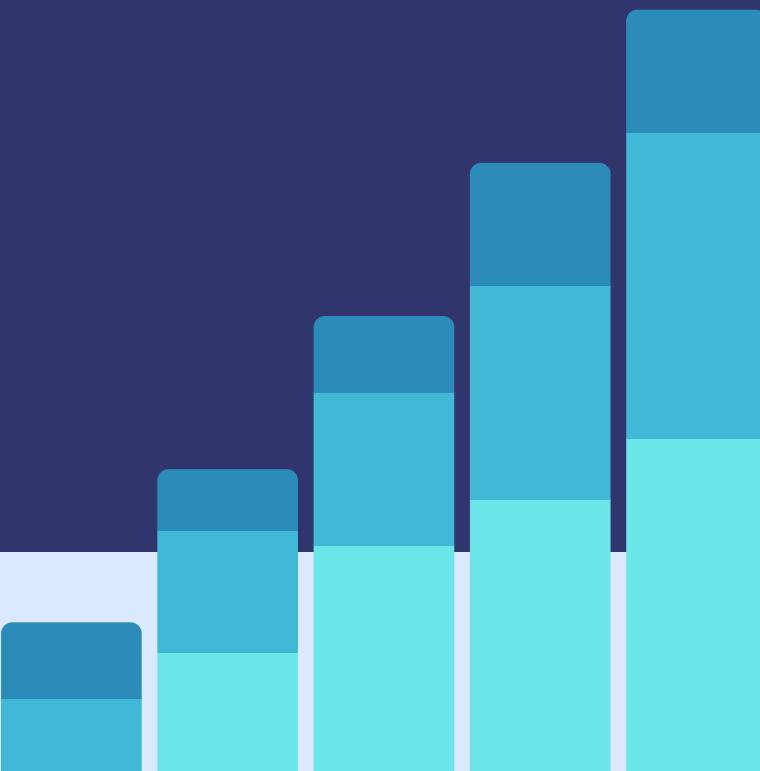
# Código de acesso à turma do Classroom:



# Repositório do github do minicurso:

<https://github.com/GiBorelli/minicursoR>

# ANÁLISE EXPLORATÓRIA DE DADOS



- Estatísticas descritivas
- Média, mediana, moda, desvio padrão
- Visualização de dados
- Histogramas, boxplots, scatter plots
- Limpeza de dados
- Tratamento de valores ausentes
- Transformações de dados

# NA

Em R, dados NA representam valores ausentes ou indefinidos em um dataset. Eles indicam que a informação para uma determinada posição está indisponível ou não foi registrada

	Nome	Idade	Salario	Departamento	Ano_de_Contratacao	Genero
1	Ana	28	3500	Marketing	2015	Feminino
2	Bruno	34	4800	TI	NA	masculino
3	Carla	29	NA	TI	2018	Feminino
4	Daniel	NA	5200	Financero	2010	Masculino
5	Eduarda	31	4500	Marketing	2014	Fememino
6	Felipe	25	3200	Financeiro	2019	Masculino

# NA

Existem maneiras diferentes de lidar com NAs

- Remover valores ausentes:
  - Usar `na.omit()` ou `na.exclude()` para excluir linhas que contenham NA.
- Substituir valores ausentes:
  - Substituir por valores plausíveis (deve-se ser analisado caso a caso). Pode ser por média, mediana, 0, manualmente.
- Ignorar durante cálculos.

# Estatística

Diversas funções estatísticas estão disponíveis em R.  
No material do curso está disponibilizado um link que mostra  
algumas delas

# Limpeza de dados

Muitas vezes é necessário fazer uma limpeza nos dados que temos.  
A limpeza pode ter que ser feita de muitas maneiras diferentes.

	<b>id</b>	<b>nome</b>	<b>idade</b>	<b>departamento</b>	<b>salario</b>	<b>data_admissao</b>	<b>cidade</b>
1	1	Maria Silva	35.0	Vendas	4500.5	2015-06-22	Sao Paulo
2	2	João Souza	28.0	Tecnologia	5300.0	2019-03-12	São paulo
3	3	Pedro Lima	45.0	RH	5200.75	2011-08-05	Rio de Janeiro
4	4	Ana Costa		Financeiro	4800.0	2018-01-18	RiodeJaneiro
5	5	Lucas Pereira	32.0	Vendas		2017-05-27	Sao Paulo

# Limpeza de dados

Diferenciação entre letras maiúsculas e minúsculas, acentos, espaços no meio das palavras, etc.

Também é importante manter as colunas no tipo certo.  
Uma coluna numérica não pode ter palavras nela, e etc.

# Limpeza de dados

Um exemplo de limpeza simples que pode ser necessária para valores numéricos.

## Arredondamento

```
arred <- 15.59
round(arred)    #16
trunc(arred)   #15
ceiling(arred)  #16
```

# Prática

- Carregue o arquivo  
<https://raw.githubusercontent.com/GiBorelli/minicursoR/refs/heads/main/datasets/vendas.csv>
- Arrume a coluna de Data
- Faça um gráfico de linhas da quantidade de cada produto vendida
- Gráfico de linha da quantidade total de produtos vendidos por dia

```
df$Data <- as.Date(df$Data, format = "%Y-%d-%m")

plot(x = df$Data[which(df$Produto=="Produto A")],
      y = df$Vendas[which(df$Produto=="Produto A")],
      type = 'l', ylim = c(50,200))
lines(x = df$Data[which(df$Produto=="Produto B")],
      y = df$Vendas [which(df$Produto=="Produto B")], col = 'red')
lines(x = df$Data[which(df$Produto=="Produto C")],
      y = df$Vendas [which(df$Produto=="Produto C")], col = 'blue')
```

```
library(dplyr)

vendas_por_dia <- df %>%
  group_by(Data) %>%
  summarise(Vendas_Totais = sum(vendas))

plot(vendas_por_dia, type ='l')
```

# type

Ao criar um gráfico por meio da função `plot()`, pode-se passar o atributo **type = “letra”**.

Na tabela estão alguns desses atributos.

Tipo de Gráfico	Atributo
Pontos	p
Linhas	l
Histograma	h
Stair Steps	s
Sem números nos eixos	n

Referência:

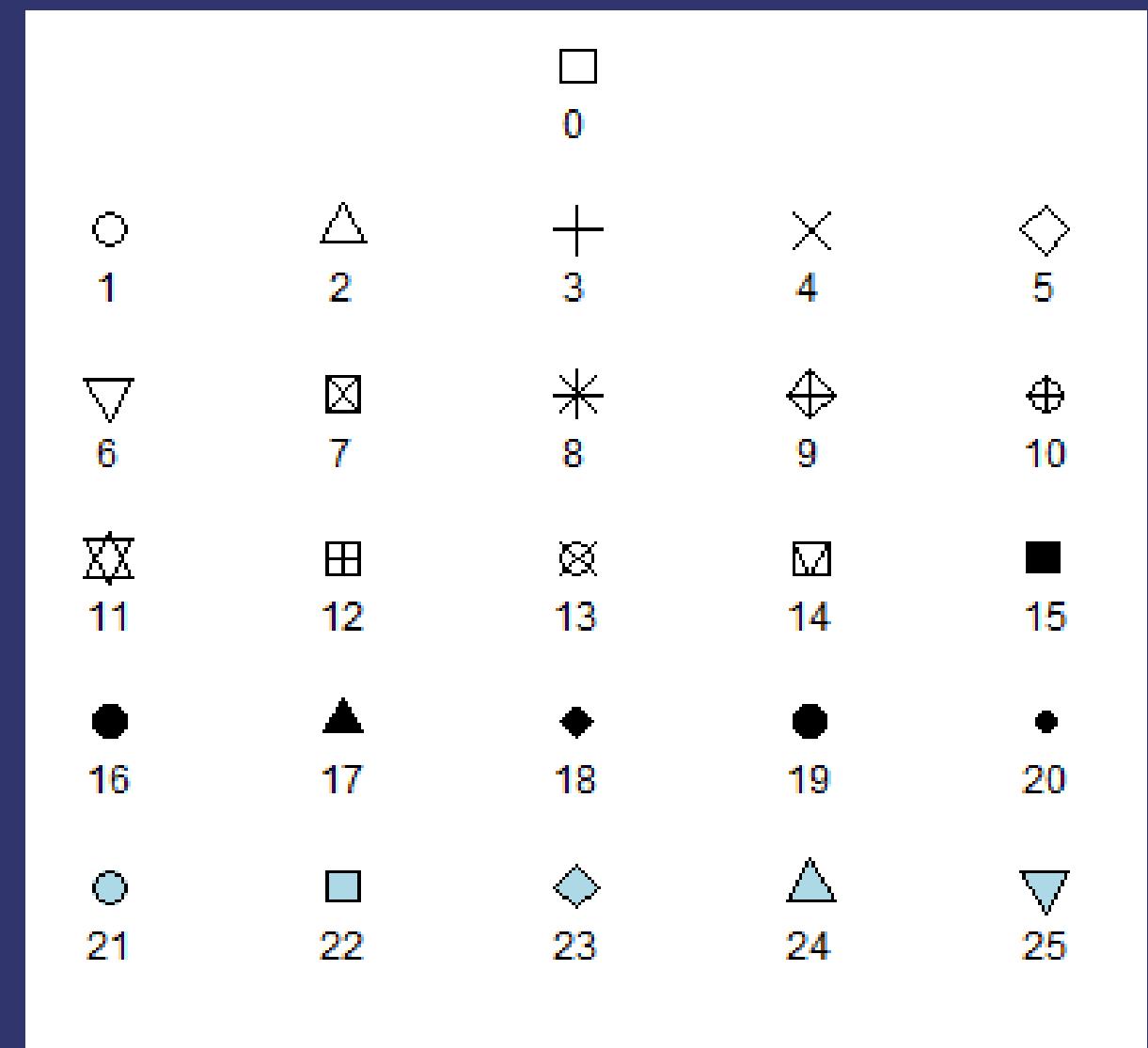
<https://www.rdocumentation.org/packages/graphics VERSIONS/3.6.2/topics/plot>

# pch

Ao criar um gráfico por meio da função `plot(type = "p")` ou `points()`, pode-se passar o atributo

**`pch = "num"`.**

Na imagem estão alguns desses atributos.

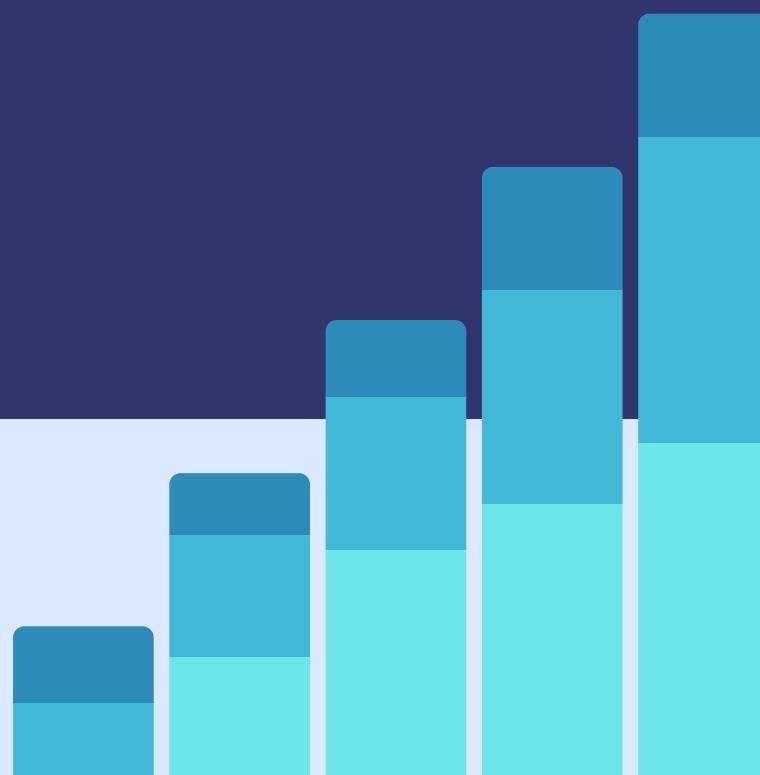


# Gráficos

Indicação de Material para Criação de outros Gráficos:

[https://vanderleidebastiani.github.io/tutoriais/Graficos\\_com\\_R.html](https://vanderleidebastiani.github.io/tutoriais/Graficos_com_R.html)

# VISUALIZAÇÃO DE DADOS AVANÇADA



- Plotagem com ggplot2
- Gráficos de linhas, barras, e outros
- Customização de gráficos (títulos, labels, temas)
- Criação de relatórios
- Introdução ao RMarkdown
- Geração de relatórios dinâmicos

# POR QUE ELES SÃO ÚTEIS?

ORGANIZAR  
INFORMAÇÕES

COMPARAR  
DADOS

FÁCIL DE  
ENTENDER E  
COMPARTILHAR

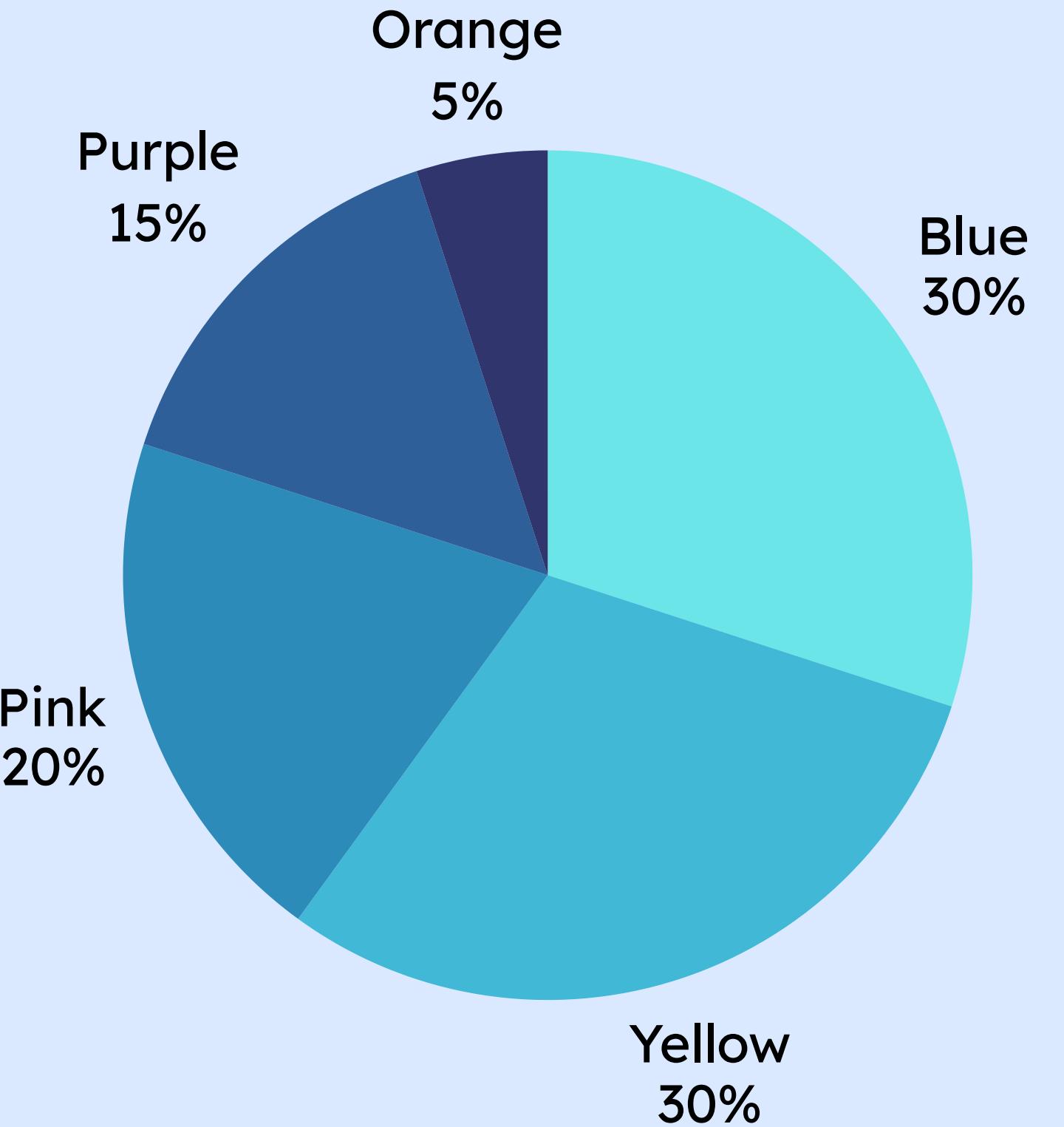


# TABELAS

A table is a grid with rows and columns for organising data/information.

Use a table when you need to organise detailed numerical or categorical information.

Subject	Number of Students
History	19
Maths	27
English	25
PE	18
Science	28



# GRÁFICO DE PIZZA

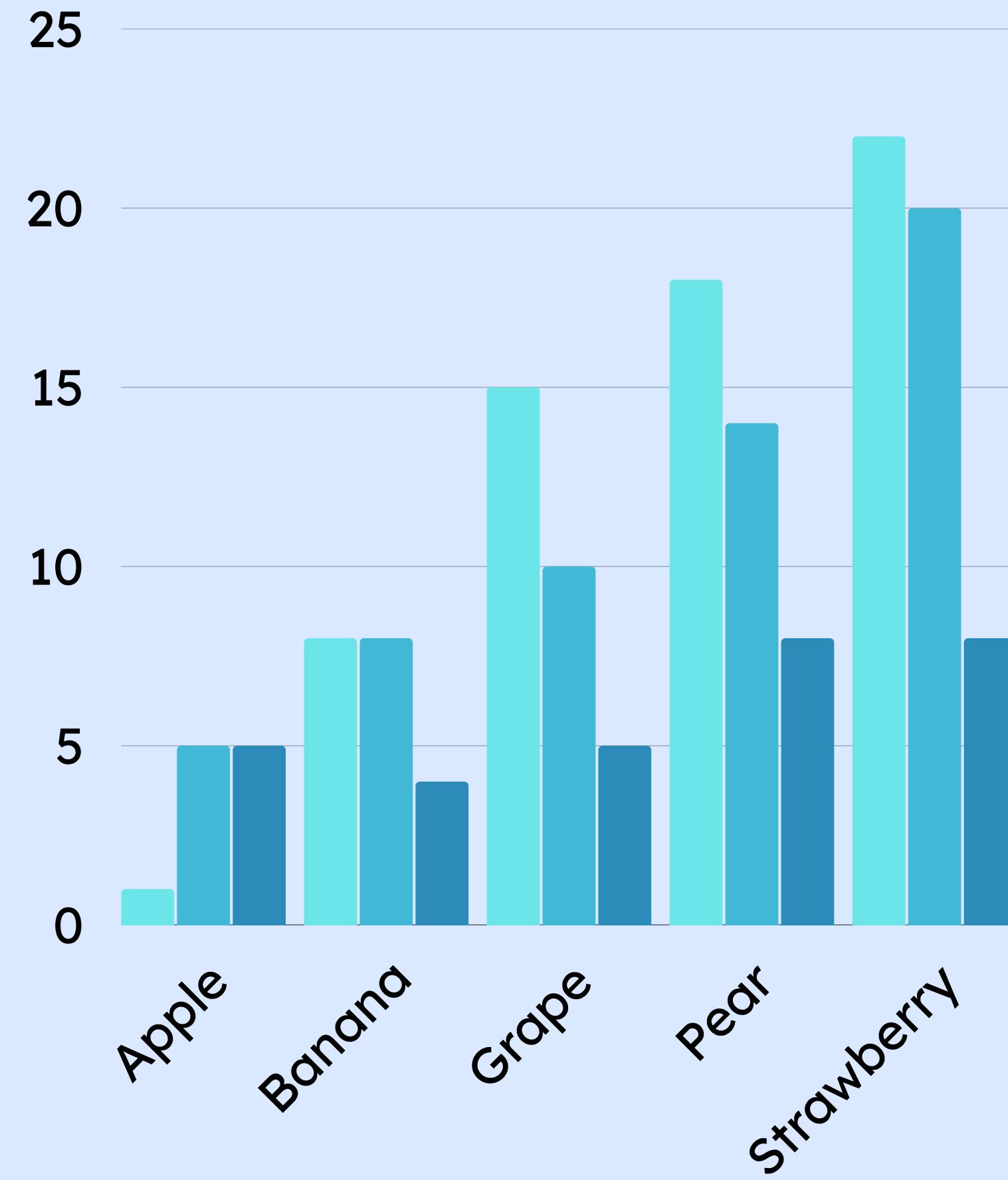
A pie chart looks like a pizza and divides up one whole into parts.

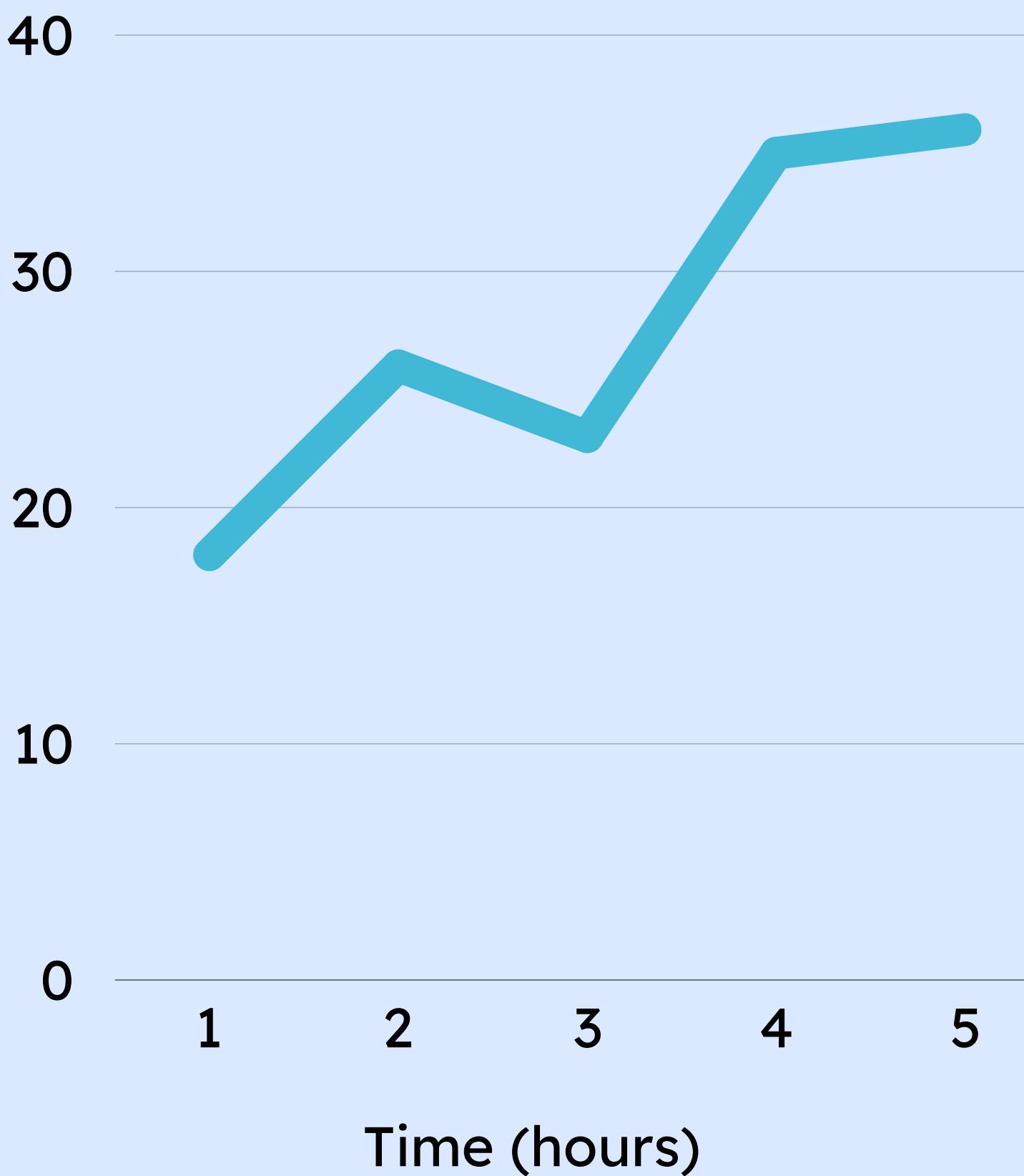
Use a pie chart when you want to show how something is divided into parts. They are great for data given in percentages!

# GRÁFICO DE BARRAS

A bar chart uses bars (vertically or horizontally) to show and compare data.

Use a bar chart when you want to compare different categories or items.





# GRÁFICO DE LINHA

A line graph uses lines to show how data changes over time.

Use a line graph when you want to track how something changes over different timescales (minutes, hour, days etc.)

# HISTOGRAMA

A histogram is like a bar chart but for numbers in ranges (not categories).

Use a histogram when you have data in different ranges, like measuring the heights of people in your class.

