

Population?

Refers to the total amount of things.

Sample Size?

Total amount of things in a sample.

Sample?

Small part of population that is used for study.

Variable?

What we are studying. (Measurable, Countable, Categorized)

Age, Age, ...

feature / column

height / weight / hair

Male, female, ...

only new

Phone Numbers

21 g car

Age

Pin code

220065
441129

$P_1 + P_2$?

Screen

Add
 $P_1 + P_2$

Score

of Pets

high

high

high

Age

33

4T

2+

MEASURE

VARIABLE

342212

Neurology

ID

Phonetic

Artwork

QUANTITATIVE DATA

DATA THAT IS MEASURED IN NUMBERS. IT DEALS WITH NUMBERS THAT MAKE SENSE TO PERFORM ARITHMETIC CALCULATIONS WITH

$+$, $-$, $/$, $*$, \sqrt{x} , $\frac{1}{x}$

QUANTITATIVE VARIABLES

HEIGHT

WEIGHT

MIDTERM SCORE

countable measure

CATEGORICAL DATA

REFERS TO THE VALUES THAT PLACE "THINGS" INTO DIFFERENT GROUPS OR CATEGORIES

CATEGORICAL VARIABLES

HAIR COLOUR

TYPE OF CAT

LETTER GRADE

Artwork

Spoken - French

used / use

Spain / Australia

Ten, Pa, Four

2 more

2 lions

green - feet

~~Star~~

Personal Information Identifier
L M K X

can

Personal Information Identifier

L M K X

Add
 $P_1 + P_2$

Score

Score

Numerical

Numbers

Vegetables

Pork

Onion

Not 5

20PC

3.0

1.2

Wheat

→ Mr

→ X

→ Aff... X

→ Aff... f

QUANTITATIVE VARIABLE

DISCRETE

REFER TO VARIABLES THAT CAN ONLY BE MEASURED IN CERTAIN NUMBERS

CONTINUOUS

REFER TO VARIABLES THAT CAN TAKE ON ANY NUMERICAL VALUE

EX: NUMBER OF PETS YOU OWN

0 1 2 30 2.7

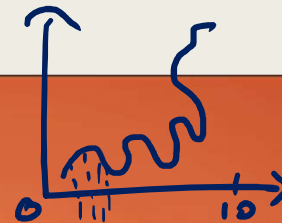
EX:

WEIGHT

105 185 170.683

1, 3, 5, 7, 4, 20

Rating 0 1 2 3 4 5



0.5 9.5
0.529 0.61

21
3
116

1.1123

1.5, 2, 2.5, 3, 3.5

1.5
1.4

1x 1.25x 1.5x 1.75x 2x

1.37

2.7

3.5

0 10
1 3

178.34

135.12
30.732

175
70.330456

CATEGORICAL VARIABLE

CATEGORICAL AND ORDINAL

LOGICAL ORDERING TO THE VALUES OF A CATEGORICAL VARIABLE

EX: LETTER GRADE

F C C+ B B+ A A+

CATEGORICAL AND NOMINAL

NO LOGICAL ORDERING TO THE VALUES OF A CATEGORICAL VARIABLE

EX: HAIR COLOUR

RED BLONDE BROWN BLUE

Formula + No logical ordering

note

1
2
3
4
5
6

Occurrence In the

City type

Tier I
Tier II
Tier III
...

City Name

1
2
3
4

Order

A+
A
B+
B
C+
D
F

Classification

I
II
III
IV
V

HDI

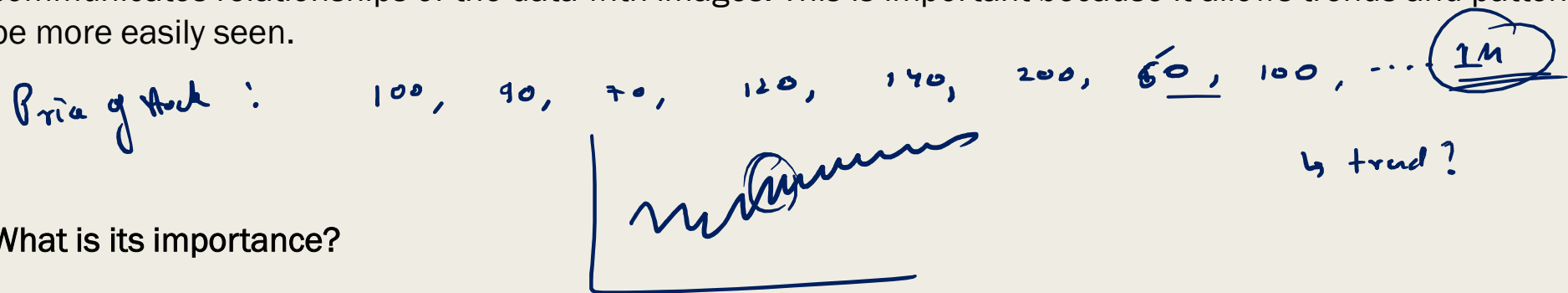
Low
Medium
High

Data Visualization Basics

What is data Visualization?

Imagine

Data visualization is the representation of data or information in a graph, chart, or other visual format. It communicates relationships of the data with images. This is important because it allows trends and patterns to be more easily seen.



What is its importance?

We need data visualization because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows on a spreadsheet. It's the way the human brain works.

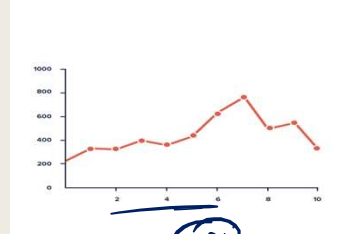
Since the purpose of data analysis is to gain insights, data is much more valuable when it is visualized.

Even if a data analyst can pull insights from data without visualization, it will be more difficult to communicate the meaning without visualization.

Charts and graphs make communicating data findings easier even if you can identify the patterns without them.

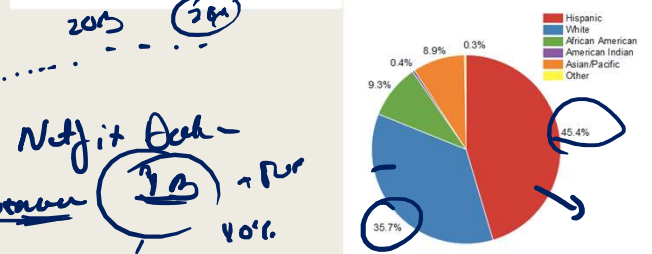
Line Chart. → 2 Variable → Time-series (Period) → 2018, 2019, ... → (X)
 → Numerical value (Y)

A line chart is, as one can imagine, a line or multiple lines showing how single, or multiple variables develop over time.



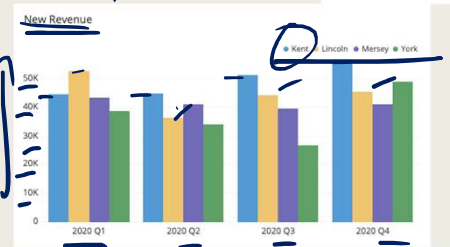
Pie Chart. → 1 Variable → Categorical Data → Distribution of your categories

A pie chart is a circular graph divided into slices. The larger a slice is the bigger portion of the total quantity it represents.

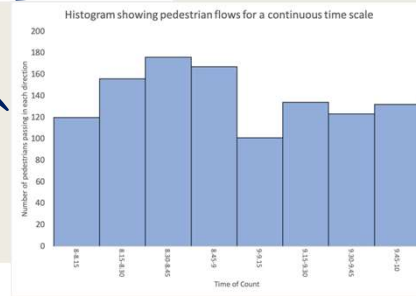
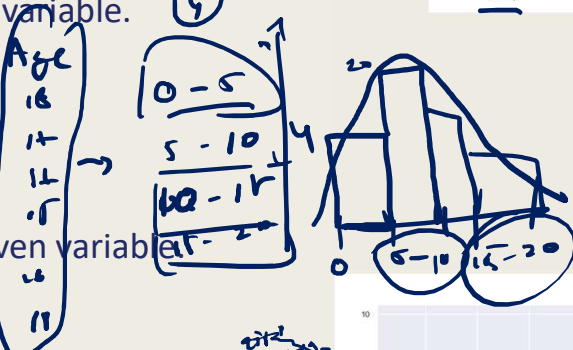


Bar Graph. → Numerical (Y-axis) → Categorical (X-axis) → Vertical bar graph

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. Can be of one variable or many variable.



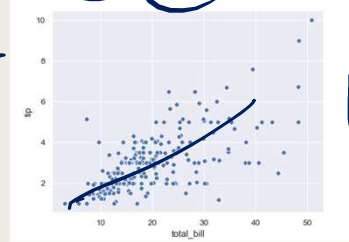
Histogram → one variable → you → frequency → x-axis → Bin (Range) → Numerical value → The spread of the data



Scatter Plots

A scatter plot is a great indicator that allows us to see whether there is a pattern to be found between two variables. E.g. : Positive or negative relationship. No relationship

2 Numerical data



HISTOGRAM

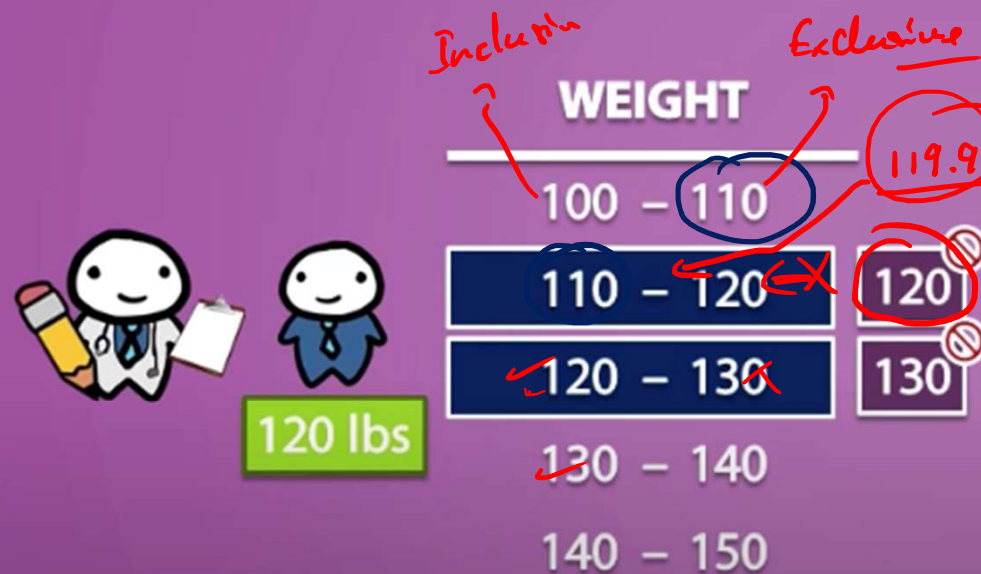


FREQUENCY DISTRIBUTION

WEIGHT		FREQUENCY
100 – 110	←	8
110 – 120	←	16
120 – 130	←	12
130 – 140	←	8
140 – 150	←	6

101, 102, 101, 101, 101
109, 109

FREQUENCY DISTRIBUTION



BY CONVENTION, WE SAY THAT EACH INTERVAL DOES NOT INCLUDE THE RIGHT END POINT

Sheer man.

50 for 100 to 110
0-10000
low I

0-5 5-10 10-15 15-20 20-25 25-30 30-35 35-40 40-45 45-50
0 0 0 1 2 3 4 5 6 7
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40
41 42 43 44 45 46 47 48 49 50
51 52 53 54 55 56 57 58 59 60
61 62 63 64 65 66 67 68 69 70
71 72 73 74 75 76 77 78 79 80
81 82 83 84 85 86 87 88 89 90
91 92 93 94 95 96 97 98 99 100
101 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120
121 122 123 124 125 126 127 128 129 130
131 132 133 134 135 136 137 138 139 140
141 142 143 144 145 146 147 148 149 150
151 152 153 154 155 156 157 158 159 160
161 162 163 164 165 166 167 168 169 170
171 172 173 174 175 176 177 178 179 180
181 182 183 184 185 186 187 188 189 190
191 192 193 194 195 196 197 198 199 200
201 202 203 204 205 206 207 208 209 210
211 212 213 214 215 216 217 218 219 220
221 222 223 224 225 226 227 228 229 230
231 232 233 234 235 236 237 238 239 240
241 242 243 244 245 246 247 248 249 250
251 252 253 254 255 256 257 258 259 260
261 262 263 264 265 266 267 268 269 270
271 272 273 274 275 276 277 278 279 280
281 282 283 284 285 286 287 288 289 290
291 292 293 294 295 296 297 298 299 300
301 302 303 304 305 306 307 308 309 310
311 312 313 314 315 316 317 318 319 320
321 322 323 324 325 326 327 328 329 330
331 332 333 334 335 336 337 338 339 340
341 342 343 344 345 346 347 348 349 350
351 352 353 354 355 356 357 358 359 360
361 362 363 364 365 366 367 368 369 370
371 372 373 374 375 376 377 378 379 380
381 382 383 384 385 386 387 388 389 390
391 392 393 394 395 396 397 398 399 400
401 402 403 404 405 406 407 408 409 410
411 412 413 414 415 416 417 418 419 420
421 422 423 424 425 426 427 428 429 430
431 432 433 434 435 436 437 438 439 440
441 442 443 444 445 446 447 448 449 450
451 452 453 454 455 456 457 458 459 460
461 462 463 464 465 466 467 468 469 470
471 472 473 474 475 476 477 478 479 480
481 482 483 484 485 486 487 488 489 490
491 492 493 494 495 496 497 498 499 500
501 502 503 504 505 506 507 508 509 510
511 512 513 514 515 516 517 518 519 520
521 522 523 524 525 526 527 528 529 530
531 532 533 534 535 536 537 538 539 540
541 542 543 544 545 546 547 548 549 550
551 552 553 554 555 556 557 558 559 560
561 562 563 564 565 566 567 568 569 570
571 572 573 574 575 576 577 578 579 580
581 582 583 584 585 586 587 588 589 590
591 592 593 594 595 596 597 598 599 600
601 602 603 604 605 606 607 608 609 610
611 612 613 614 615 616 617 618 619 620
621 622 623 624 625 626 627 628 629 630
631 632 633 634 635 636 637 638 639 640
641 642 643 644 645 646 647 648 649 650
651 652 653 654 655 656 657 658 659 660
661 662 663 664 665 666 667 668 669 670
671 672 673 674 675 676 677 678 679 680
681 682 683 684 685 686 687 688 689 690
691 692 693 694 695 696 697 698 699 700
701 702 703 704 705 706 707 708 709 710
711 712 713 714 715 716 717 718 719 720
721 722 723 724 725 726 727 728 729 730
731 732 733 734 735 736 737 738 739 740
741 742 743 744 745 746 747 748 749 750
751 752 753 754 755 756 757 758 759 760
761 762 763 764 765 766 767 768 769 770
771 772 773 774 775 776 777 778 779 780
781 782 783 784 785 786 787 788 789 790
791 792 793 794 795 796 797 798 799 800
801 802 803 804 805 806 807 808 809 810
811 812 813 814 815 816 817 818 819 820
821 822 823 824 825 826 827 828 829 830
831 832 833 834 835 836 837 838 839 840
841 842 843 844 845 846 847 848 849 850
851 852 853 854 855 856 857 858 859 860
861 862 863 864 865 866 867 868 869 870
871 872 873 874 875 876 877 878 879 880
881 882 883 884 885 886 887 888 889 890
891 892 893 894 895 896 897 898 899 900
901 902 903 904 905 906 907 908 909 910
911 912 913 914 915 916 917 918 919 920
921 922 923 924 925 926 927 928 929 930
931 932 933 934 935 936 937 938 939 940
941 942 943 944 945 946 947 948 949 950
951 952 953 954 955 956 957 958 959 960
961 962 963 964 965 966 967 968 969 970
971 972 973 974 975 976 977 978 979 980
981 982 983 984 985 986 987 988 989 990
991 992 993 994 995 996 997 998 999 1000

Class Interval

100 lower
salary
210,000 min
210,000 max

WEIGHT	FREQUENCY	CALCULATIONS	RELATIVE FREQUENCY
100 - 110	8	$8 \div 50 =$	0.16
110 - 120	16	$16 \div 50 =$	0.32
120 - 130	12	$12 \div 50 =$	0.24
130 - 140	8	$8 \div 50 =$	0.16
140 - 150	6	$6 \div 50 =$	0.12
SUM = 50			SUM = 1

$\frac{8}{50} + \frac{16}{50} + \frac{12}{50} + \frac{8}{50} + \frac{6}{50} = 1$

Frequency \rightarrow Count of object/Value

210--



150-160

Or. 10.6%