Intro: Random Variables

Mr. Merrick · September 15, 2025

Consider the random sample from tikz village shown below. Tabulate the sample into a tidy data-frame.

Career	Weight (kg)	Height (cm)	Income ($k)	Pets
Engineer	86	183	118	1
Businessman	81	178	132	0
Chef	73	171	62	2
Nurse	66	168	94	1
Physician	78	185	228	1
Pilot	82	181	172	0
Cowboy	84	186	55	2
Builder	92	187	78	0
Businessman	88	182	149	1
Chef	69	168	57	3
Nurse	72	174	88	0
Physician	74	179	204	2
Pilot	79	176	165	1
Cowboy	77	180	51	1
Engineer	89	186	127	2
Pilot	86	184	181	1
Physician	70	177	193	0
Nurse	64	165	86	1
Chef	82	178	67	1
Businessman	76	175	121	0
Cowboy	80	182	58	3
Businessman	90	186	171	1
Builder	85	181	72	2
Nurse	69	170	95	2
Physician	82	188	239	0
Pilot	75	178	159	0
Chef	77	173	64	2
Cowboy	79	179	47	1
Physician	73	176	212	2
Nurse	67	169	90	0
Pilot	88	185	186	1
Engineer	83	180	112	0
Businessman	84	180	138	1
Chef	70	169	55	1
Cowboy	85	183	53	2

1. How many *variables* were recorded from the sample? **5** How many *observations*? **35**

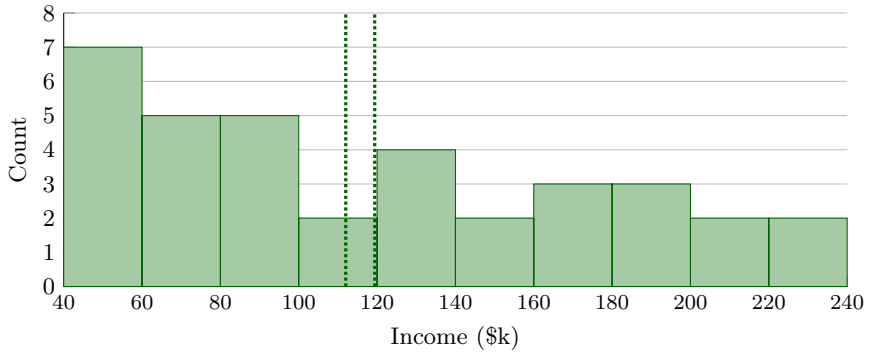
2. Classify each variable (circle one per section).

Variable	Type			Scale			
	Categorical	Quantitative — Discrete.	Quantitative — Continuous	Nominal	Ordinal	Ratio	Interval
Career	✓	◯	◯	✓	◯	◯	◯
Weight	◯	◯	✓	◯	◯	✓	◯
Height	◯	◯	✓	◯	◯	✓	◯
Income	◯	◯	✓	◯	◯	✓	◯
Pets	◯	✓	◯	◯	◯	✓	◯

3. Histogram of Income (in $k)

Tally the counts in each bin (half-open: [a,b)). Add dotted vertical lines for the sample *median* and *mean*.

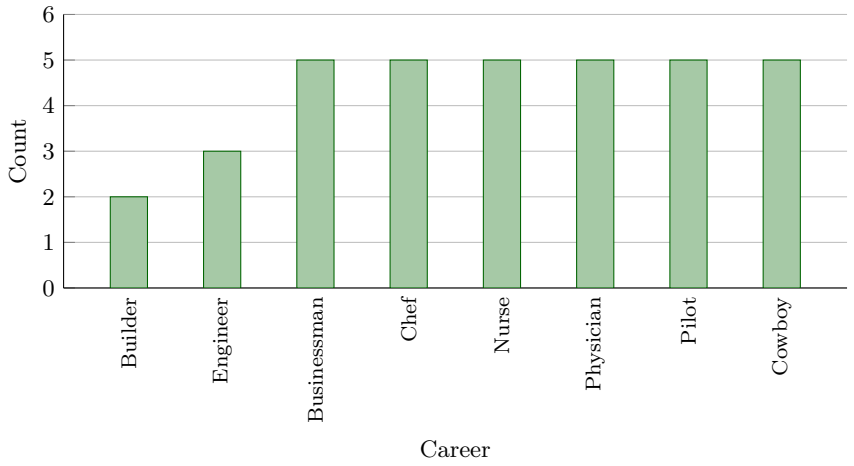
Bin	Count
[40,60)	**7**
[60,80)	**5**
[80,100)	**5**
[100,120)	**2**
[120,140)	**4**
[140,160)	**2**
[160,180)	**3**
[180,200)	**3**
[200,220)	**2**
[220,240)	**2**



4. Bar Chart of Career

Tally the counts and proportions by career, then draw the bars.

Career	Count	Proportion
Builder	**2**	**0.06**
Engineer	**3**	**0.09**
Businessman	**5**	**0.14**
Chef	**5**	**0.14**
Nurse	**5**	**0.14**
Physician	**5**	**0.14**
Pilot	**5**	**0.14**
Cowboy	**5**	**0.14**

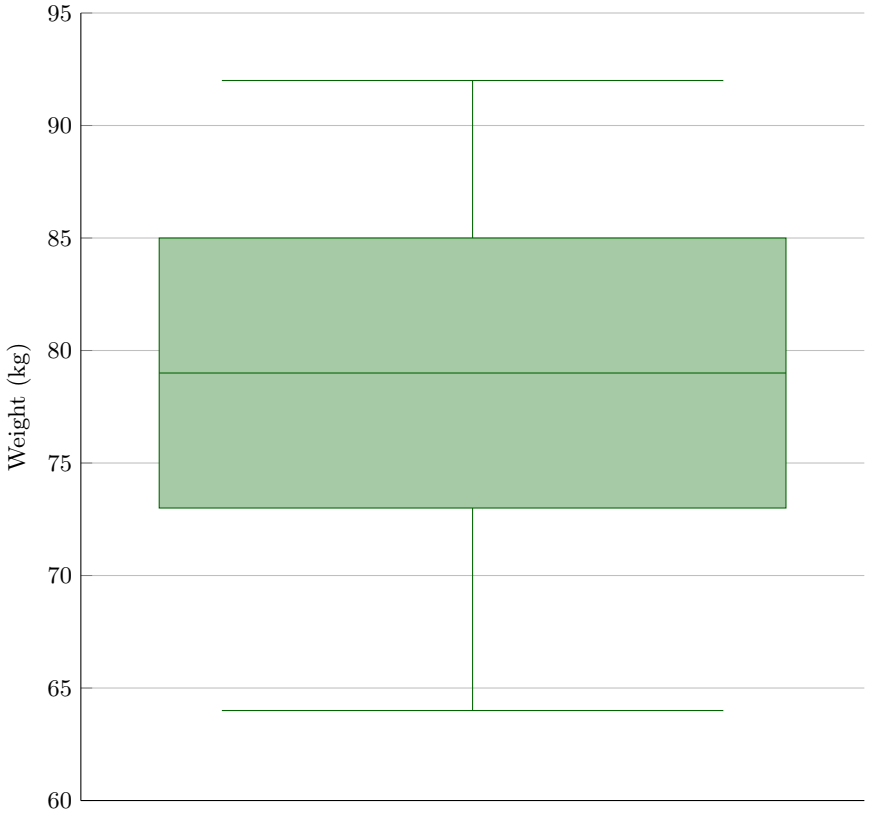


5. Box plot of Weights (kg)

Use the sorted weights below to compute the five-number summary, then sketch the box plot. In solutions mode the box is drawn automatically.

64	66	67	69	69	70	70
72	73	73	75	76	77	77
78	79	79	80	81	82	82
82	83	84	84	85	85	86
86	88	88	89	90	92	

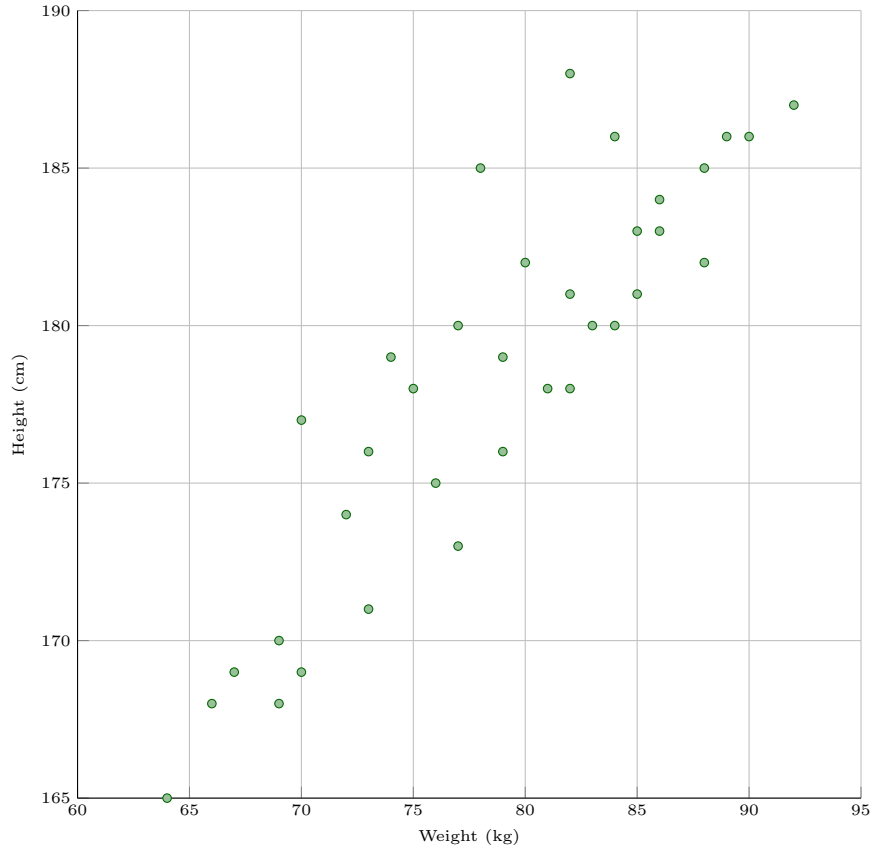
Minimum	**64**
Q1	**73**
Median (Q2)	**79**
Q3	**85**
Maximum	**92**



6. Scatter plot: Weight (kg) vs. Height (cm)

Use the tabulated pairs (kg, cm) to plot the scatter. In solutions mode the points are drawn automatically.

(W,H)		(W,H)		(W,H)		(W,H)		(W,H)	
86	183	81	178	73	171	66	168	78	185
82	181	84	186	92	187	88	182	69	168
72	174	74	179	79	176	77	180	89	186
86	184	70	177	64	165	82	178	76	175
80	182	90	186	85	181	69	170	82	188
75	178	77	173	79	179	73	176	67	169
88	185	83	180	84	180	70	169	85	183



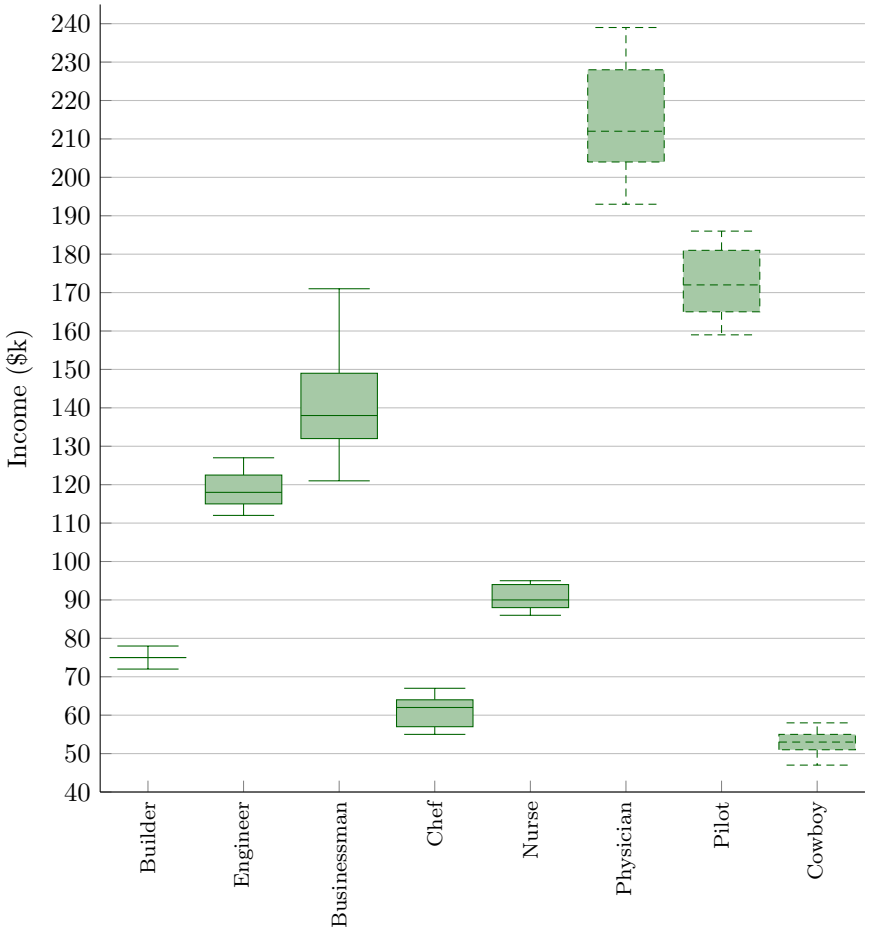
7. What does the scatter plot suggest about the relationship between *Weight* and *Height*? Briefly describe the direction and strength.

There is a positive, roughly linear association: taller individuals tend to weigh more (moderate strength).

8. Income by Career — Pairwise Box Plots

Draw pairwise boxplots so that each box summarizes the income distribution (in \$k) for a career.

Career	Min	Q1	Median	Q3	Max
Builder	72	72	75	78	78
Engineer	112	112	118	127	127
Businessman	121	132	138	149	171
Chef	55	57	62	64	67
Nurse	86	88	90	94	95
Physician	193	204	212	228	239
Pilot	159	165	172	181	186
Cowboy	47	51	53	55	58



9. Basic Probability

Let one person be selected uniformly at random from the 35 individuals. Provide answers as exact fractions and/or decimals.

1. $P(\text{Physician}) = \frac{5}{35} \approx 0.143$
2. $P(\text{Physician and Income} > 200\text{k}) = \frac{4}{35} \approx 0.114$
3. $P(\text{Income} \geq 100\text{k}) = \frac{18}{35} \approx 0.514$
4. $P(\text{Engineer} \mid \text{Income} \geq 100\text{k}) = \frac{3}{18} = \frac{1}{6} \approx 0.167$
5. $P(\text{Pilot or Physician}) = \frac{10}{35} = \frac{2}{7} \approx 0.286$
6. $P(\text{Pets} \geq 2) = \frac{14}{35} = \frac{2}{5} = 0.400$
7. $P(40\text{k} \leq \text{Income} < 100\text{k}) = \frac{17}{35} \approx 0.486$