Lecture 1
Introduction to Data Science

1. Introduction to Data

2. A Look Ahead

1. Introduction to Data

What Does Data Look Like?

What Does Data Look Like?

titanic

	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

- **pclass** → Indicator of children asocio-economic status (1 = highest class).
- **sibsp** → Number of siblings or spouses aboard the ship.
- **parch** → Number of parents or board the ship.
- **fare** → Ticket fare (price paid for the ticket).
- **embarked** → Port where the passenger boarded the ship.
- **boat** → The number of the lifeboat that the passenger boarded.
- **body** → if found dead body number

What Does Data Look Like?

titanic

	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Code	Port Name (English)	Country
C	Cherbourg	France
Q	Queenstown	Ireland
S	Southampton	England

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

What Does Data Look Like?

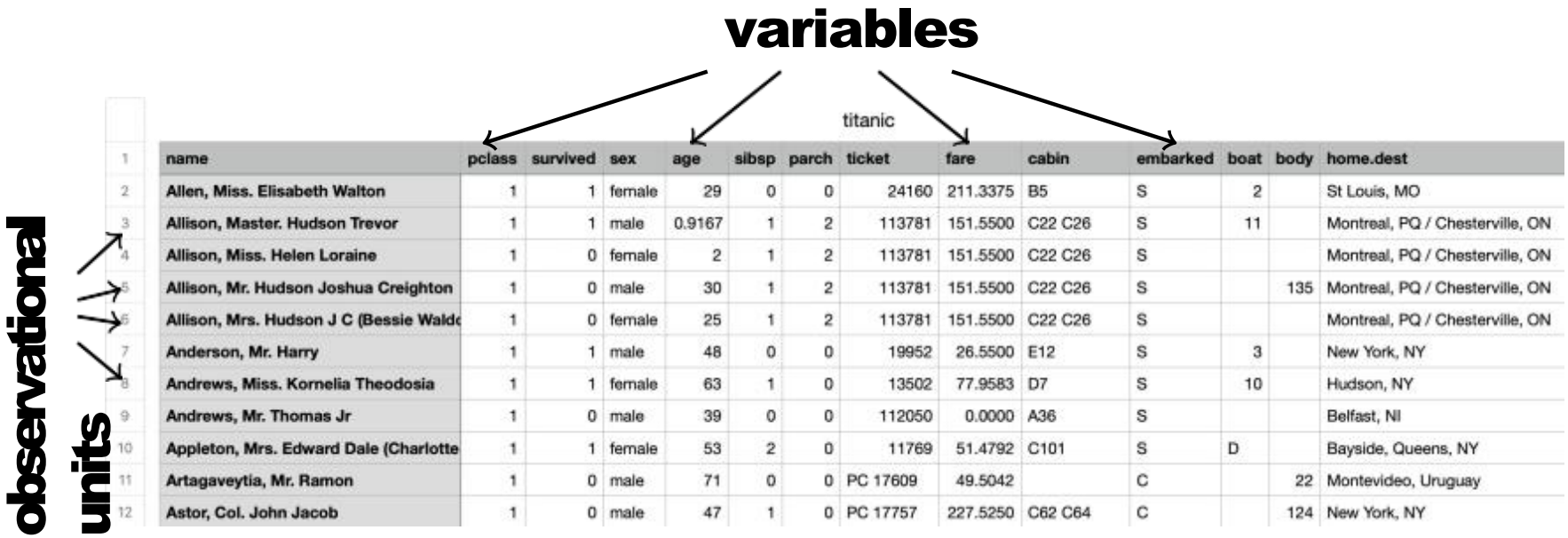
observational units

titanic

	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

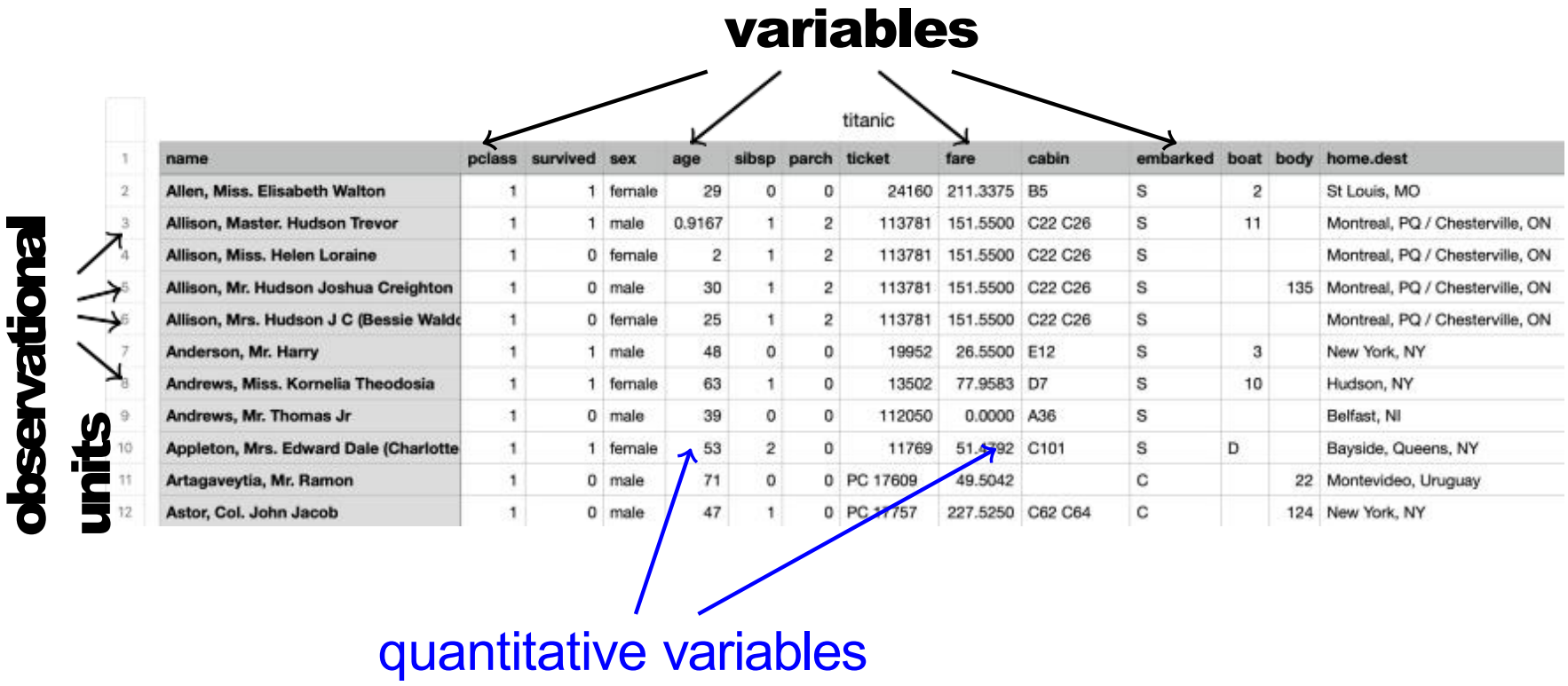
What Does Data Look Like?



	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

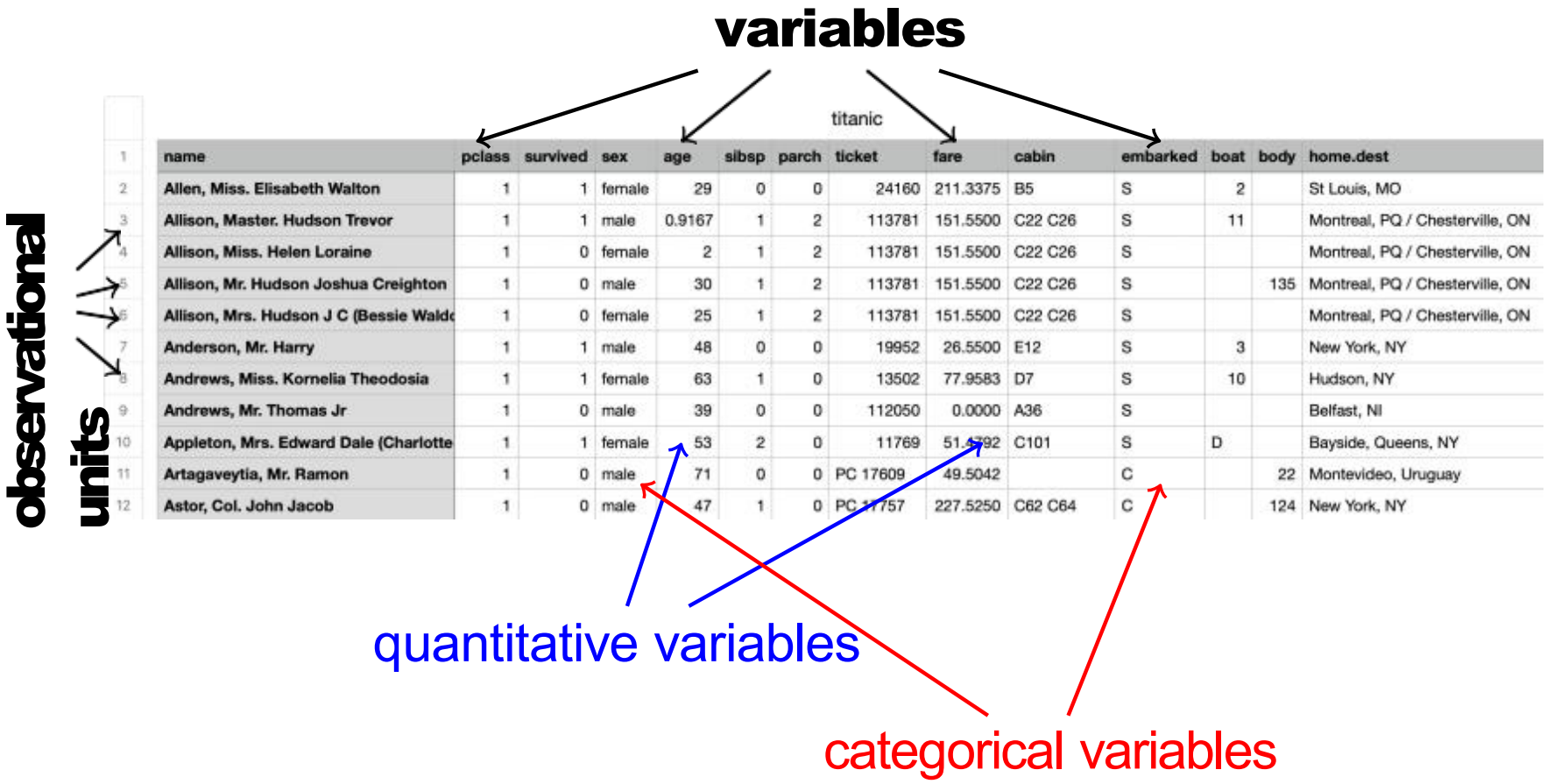
What Does Data Look Like?



	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

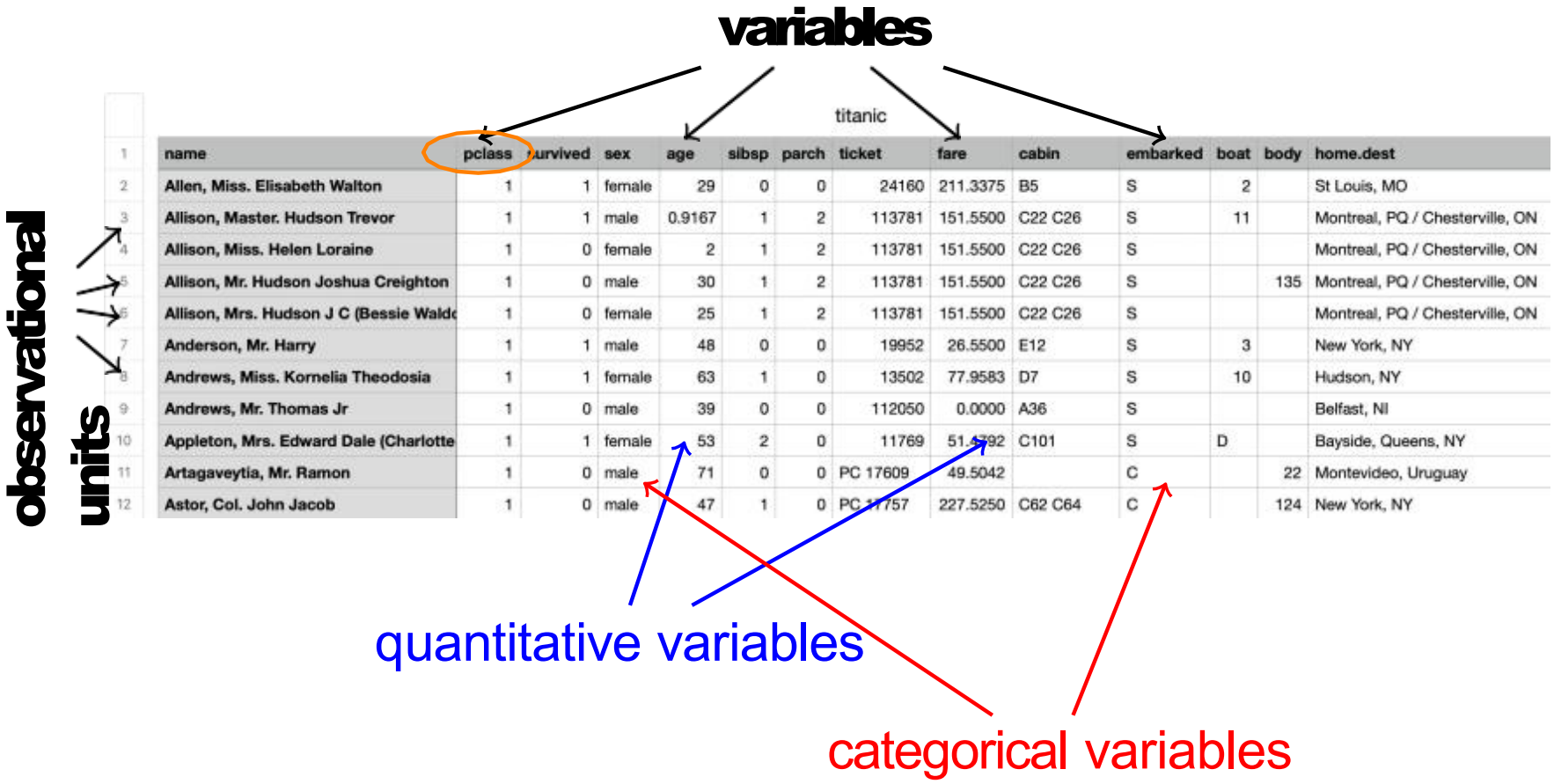
What Does Data Look Like?



	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

What Does Data Look Like?

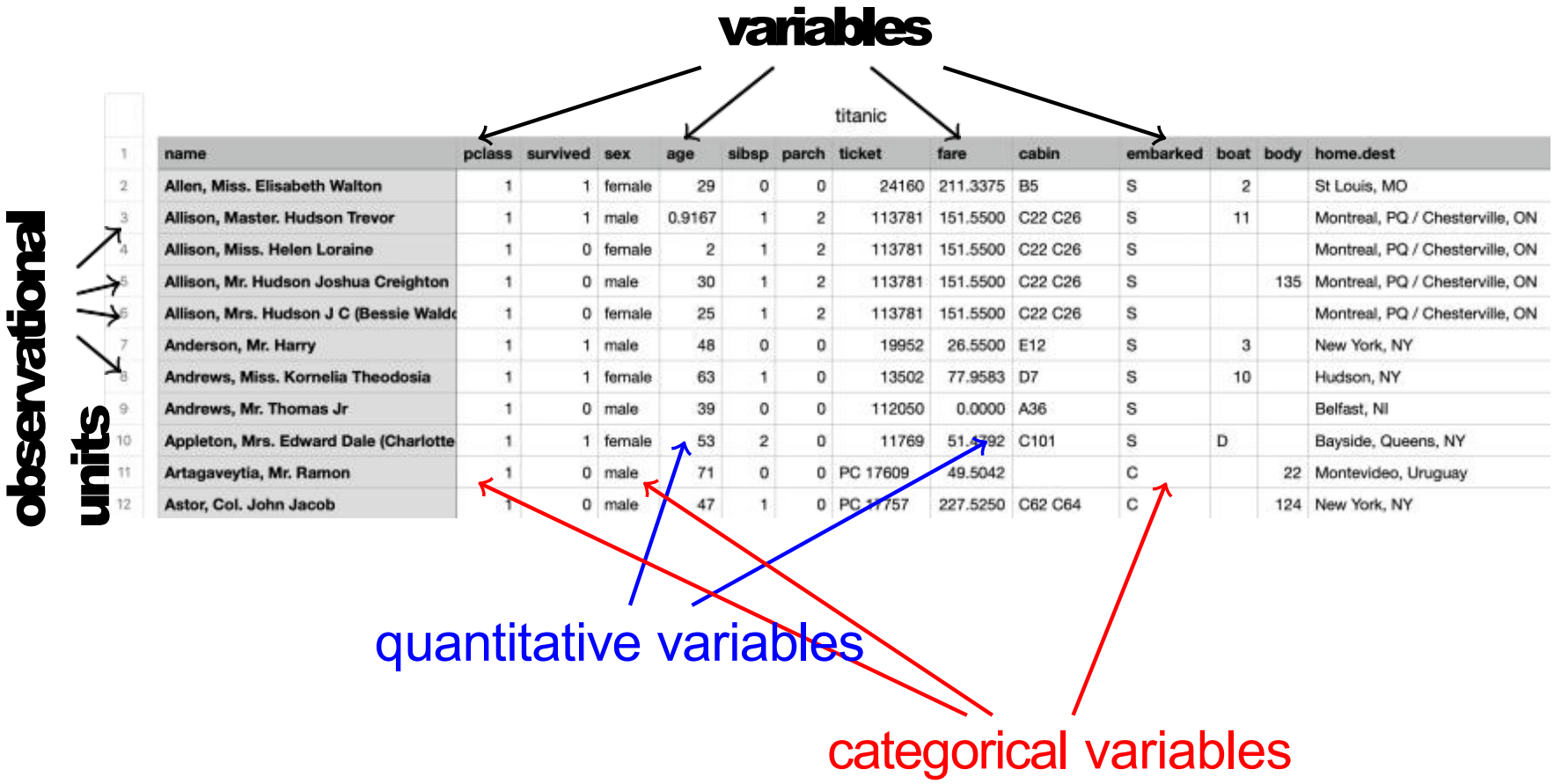


titanic

	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

What Does Data Look Like?



titanic

	name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
5	Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ / Chesterville, ON
7	Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
8	Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
10	Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

Data like this, that can be stored in a spreadsheet, is called **tabular data**.

How is Tabular Data Represented on Disk?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

How is Tabular Data Represented on Disk?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

```
name,pclass,survived,sex,age,sibsp,parch,ticket,fare,cabin,embarked,boat,body,home.dest
"Allen, Miss. Elisabeth Walton",1,1,female,29,0,0,24160,211.3375,B5,S,2,,"St Louis, MO"
"Allison, Master. Hudson Trevor",1,1,male,0.9167,1,2,113781,151.5500,C22 C26,S,11,,"Montreal, PQ / Cheste
"Allison, Miss. Helen Loraine",1,0,female,2,1,2,113781,151.5500,C22 C26,S,,,"Montreal, PQ / Chesterville,
"Allison, Mr. Hudson Joshua Creighton",1,0,male,30,1,2,113781,151.5500,C22 C26,S,,135,"Montreal, PQ / Che
"Allison, Mrs. Hudson J C (Bessie Waldo Daniels)",1,0,female,25,1,2,113781,151.5500,C22 C26,S,,,"Montreal
"Anderson, Mr. Harry",1,1,male,48,0,0,19952,26.5500,E12,S,3,,"New York, NY"
"Andrews, Miss. Kornelia Theodosia",1,1,female,63,1,0,13502,77.9583,D7,S,10,,"Hudson, NY"
"Andrews, Mr. Thomas Jr",1,0,male,39,0,0,112050,0.0000,A36,S,,,"Belfast, NI"
"Appleton, Mrs. Edward Dale (Charlotte Lamson)",1,1,female,53,2,0,11769,51.4792,C101,S,D,,"Bayside, Queen
"Artagaveytia, Mr. Ramon",1,0,male,71,0,0,PC 17609,49.5042,,C,,22,"Montevideo, Uruguay"
"Astor, Col. John Jacob",1,0,male,47,1,0,PC 17757,227.5250,C62 C64,C,,124,"New York, NY"
```

How is Tabular Data Represented on Disk?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

```
name,pclass,survived,sex,age,sibsp,parch,ticket,fare,cabin,embarked,boat,body,home.dest
"Allen, Miss. Elisabeth Walton",1,1,female,29,0,0,24160,211.3375,B5,S,2,,"St Louis, MO"
"Allison, Master. Hudson Trevor",1,1,male,0.9167,1,2,113781,151.5500,C22 C26,S,11,,"Montreal, PQ / Cheste
"Allison, Miss. Helen Loraine",1,0,female,2,1,2,113781,151.5500,C22 C26,S,,,"Montreal, PQ / Chesterville,
"Allison, Mr. Hudson Joshua Creighton",1,0,male,30,1,2,113781,151.5500,C22 C26,S,,135,"Montreal, PQ / Che
"Allison, Mrs. Hudson J C (Bessie Waldo Daniels)",1,0,female,25,1,2,113781,151.5500,C22 C26,S,,,"Montreal
"Anderson, Mr. Harry",1,1,male,48,0,0,19952,26.5500,E12,S,3,,"New York, NY"
"Andrews, Miss. Kornelia Theodosia",1,1,female,63,1,0,13502,77.9583,D7,S,10,,"Hudson, NY"
"Andrews, Mr. Thomas Jr",1,0,male,39,0,0,112050,0.0000,A36,S,,,"Belfast, NI"
"Appleton, Mrs. Edward Dale (Charlotte Lamson)",1,1,female,53,2,0,11769,51.4792,C101,S,D,,"Bayside, Queen
"Artagaveytia, Mr. Ramon",1,0,male,71,0,0,PC 17609,49.5042,,C,,22,"Montevideo, Uruguay"
"Astor, Col. John Jacob",1,0,male,47,1,0,PC 17757,227.5250,C62 C64,C,,124,"New York, NY"
```

Comma-Separated Values (CSV) format

How is Tabular Data Represented in Python?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

How is Tabular Data Represented in Python?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

DataFrame

How is Tabular Data Represented in Python?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

DataFrame

Let's interact with this data using Python in a **notebook**.

How is Tabular Data Represented in Python?

name	pclass	survived	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
Allen, Miss. Elisabeth Walton	1	1	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
Allison, Master. Hudson Trevor	1	1	male	0.9167	1	2	113781	151.5500	C22 C26	S	11		Montreal, PQ
Allison, Miss. Helen Loraine	1	0	female	2	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Allison, Mr. Hudson Joshua Creighton	1	0	male	30	1	2	113781	151.5500	C22 C26	S		135	Montreal, PQ
Allison, Mrs. Hudson J C (Bessie Waldo	1	0	female	25	1	2	113781	151.5500	C22 C26	S			Montreal, PQ
Anderson, Mr. Harry	1	1	male	48	0	0	19952	26.5500	E12	S	3		New York, NY
Andrews, Miss. Kornelia Theodosia	1	1	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
Andrews, Mr. Thomas Jr	1	0	male	39	0	0	112050	0.0000	A36	S			Belfast, NI
Appleton, Mrs. Edward Dale (Charlotte	1	1	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Qu
Artagaveytia, Mr. Ramon	1	0	male	71	0	0	PC 17609	49.5042		C		22	Montevideo,
Astor, Col. John Jacob	1	0	male	47	1	0	PC 17757	227.5250	C62 C64	C		124	New York, NY

DataFrame

Let's interact with this data using Python in a **notebook**.

All of our code will be written in Jupyter notebooks like this one.

Review: Categorical Variables

To *summarize* a categorical variable, we report the **counts** of each possible category.

Review: Categorical Variables

To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

Review: Categorical Variables

To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3    709
1    323
2    277
Name: pclass, dtype: int64
```

Review: Categorical Variables

To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3    709
1    323
2    277
Name: pclass, dtype: int64
```

To *visualize* a categorical variable, we make a **bar plot**.

```
df["pclass"].value_counts().plot.bar()
```

Review: Categorical Variables

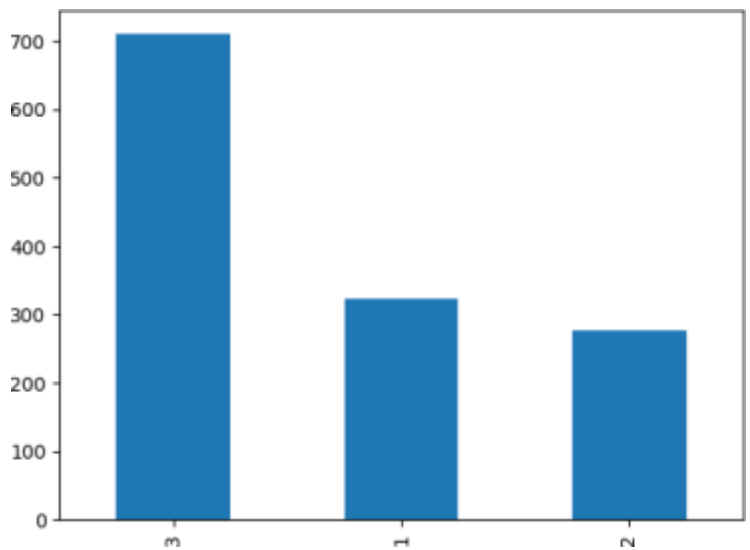
To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3      709
1      323
2      277
Name: pclass, dtype: int64
```

To *visualize* a categorical variable, we make a **bar plot**.

```
df["pclass"].value_counts().plot.bar()
```



Review: Categorical Variables

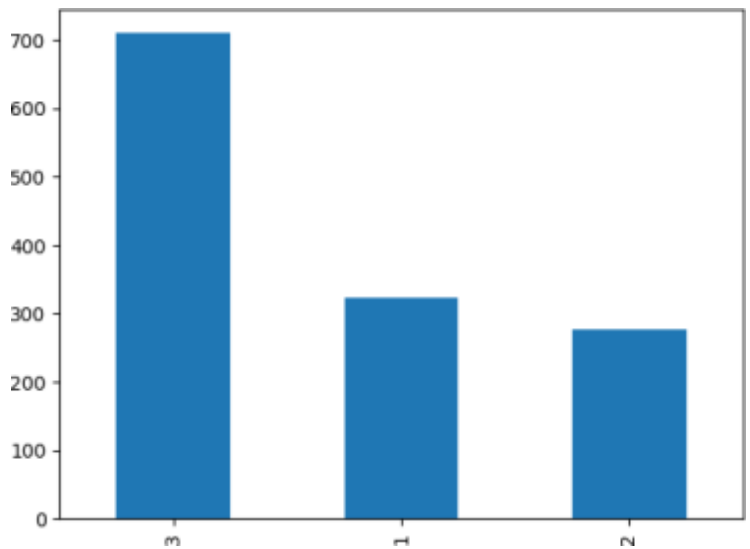
To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3      709
1      323
2      277
Name: pclass, dtype: int64
```

To *visualize* a categorical variable, we make a **bar plot**.

```
df["pclass"].value_counts().plot.bar()
```



Hmm...why are the classes out of order?

Review: Categorical Variables

To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3    709
1    323
2    277
Name: pclass, dtype: int64
```

To *visualize* a categorical variable, we make a **bar plot**.

```
df["pclass"].value_counts().sort_index().plot.bar()
```

Review: Categorical Variables

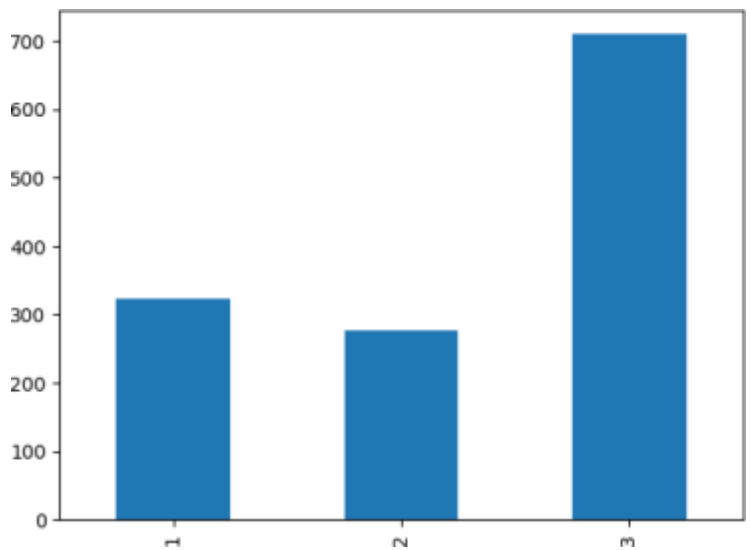
To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3    709
1    323
2    277
Name: pclass, dtype: int64
```

To *visualize* a categorical variable, we make a **bar plot**.

```
df["pclass"].value_counts().sort_index().plot.bar()
```



Review: Categorical Variables

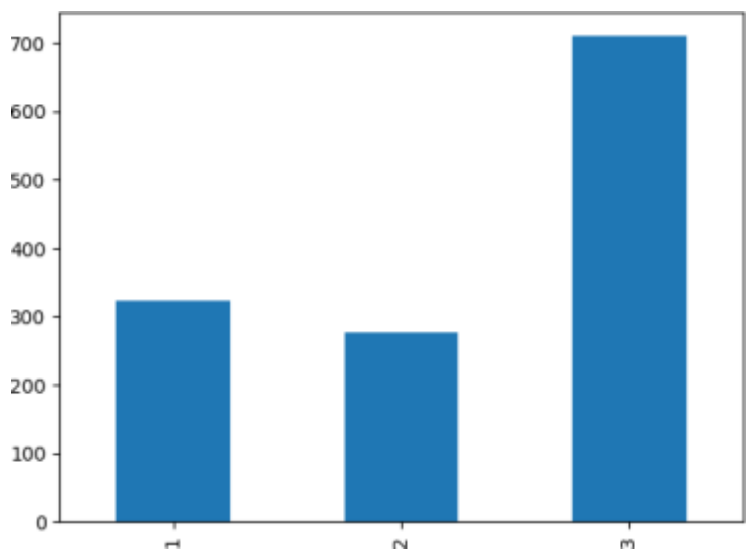
To *summarize* a categorical variable, we report the **counts** of each possible category.

```
df["pclass"].value_counts()
```

```
3    709
1    323
2    277
Name: pclass, dtype: int64
```

To *visualize* a categorical variable, we make a **bar plot**.

```
df["pclass"].value_counts().sort_index().plot.bar()
```



Notice that we can chain methods, one after the other.

1. Introduction to Data
2. A Look Ahead

The Three Parts of Data Science

1. summarizing and visualizing tabular data

The Three Parts of Data Science

1. summarizing and visualizing tabular data
2. other shapes of data: textual, hierarchical, geospatial

The Three Parts of Data Science

1. summarizing and visualizing tabular data
2. other shapes of data: textual, hierarchical, geospatial
3. machine learning