

Data Wrangling

Formats

CSV Files

Simple, universal

```
species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex
Adelie, Torgersen, 39.1, 18.7, 181, 3750, MALE
Adelie, Torgersen, 39.5, 17.4, 186, 3800, FEMALE
Adelie, Torgersen, 40.3, 18, 195, 3250, FEMALE
Adelie, Torgersen, , , , ,
Adelie, Torgersen, 36.7, 19.3, 193, 3450, FEMALE
Adelie, Torgersen, 39.3, 20.6, 190, 3650, MALE
Adelie, Torgersen, 38.9, 17.8, 181, 3625, FEMALE
Adelie, Torgersen, 39.2, 19.6, 195, 4675, MALE
Adelie, Torgersen, 34.1, 18.1, 193, 3475,
Adelie, Torgersen, 42, 20.2, 190, 4250,
Adelie, Torgersen, 37.8, 17.1, 186, 3300,
Adelie, Torgersen, 37.8, 17.3, 180, 3700,
Adelie, Torgersen, 41.1, 17.6, 182, 3200, FEMALE
Adelie, Torgersen, 38.6, 21.2, 191, 3800, MALE
Adelie, Torgersen, 34.6, 21.1, 198, 4400, MALE
Adelie, Torgersen, 36.6, 17.8, 185, 3700, FEMALE
Adelie, Torgersen, 38.7, 19, 195, 3450, FEMALE
Adelie, Torgersen, 42.5, 20.7, 197, 4500, MALE
Adelie, Torgersen, 34.4, 18.4, 184, 3325, FEMALE
Adelie, Torgersen, 46, 21.5, 194, 4200, MALE
Adelie, Biscoe, 37.8, 18.3, 174, 3400, FEMALE
Adelie, Biscoe, 37.7, 18.7, 180, 3600, MALE
Adelie, Biscoe, 35.9, 19.2, 189, 3800, FEMALE
Adelie, Biscoe, 38.2, 18.1, 185, 3950, MALE
Adelie, Biscoe, 38.8, 17.2, 180, 3800, MALE
Adelie, Biscoe, 35.3, 18.9, 187, 3800, FEMALE
Adelie, Biscoe, 40.6, 18.6, 183, 3550, MALE
Adelie, Biscoe, 40.5, 17.9, 187, 3200, FEMALE
Adelie, Biscoe, 37.9, 18.6, 172, 3150, FEMALE
```

Separate columns
with “,” *delimiter*

One observation
per line

CSV Files

Simple, universal

Column labels in first row
(*header*)

```
species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex
Adelie, Torgersen, 39.1, 18.7, 181, 3750, MALE
Adelie, Torgersen, 39.5, 17.4, 186, 3800, FEMALE
Adelie, Torgersen, 40.3, 18, 195, 3250, FEMALE
Adelie, Torgersen, , , , ,
Adelie, Torgersen, 36.7, 19.3, 193, 3450, FEMALE
Adelie, Torgersen, 39.3, 20.6, 190, 3650, MALE
Adelie, Torgersen, 38.9, 17.8, 181, 3625, FEMALE
Adelie, Torgersen, 39.2, 19.6, 195, 4675, MALE
Adelie, Torgersen, 34.1, 18.1, 193, 3475,
Adelie, Torgersen, 42, 20.2, 190, 4250,
Adelie, Torgersen, 37.8, 17.1, 186, 3300,
Adelie, Torgersen, 37.8, 17.3, 180, 3700,
Adelie, Torgersen, 41.1, 17.6, 182, 3200, FEMALE
Adelie, Torgersen, 38.6, 21.2, 191, 3800, MALE
Adelie, Torgersen, 34.6, 21.1, 198, 4400, MALE
Adelie, Torgersen, 36.6, 17.8, 185, 3700, FEMALE
Adelie, Torgersen, 38.7, 19, 195, 3450, FEMALE
Adelie, Torgersen, 42.5, 20.7, 197, 4500, MALE
Adelie, Torgersen, 34.4, 18.4, 184, 3325, FEMALE
Adelie, Torgersen, 46, 21.5, 194, 4200, MALE
Adelie, Biscoe, 37.8, 18.3, 174, 3400, FEMALE
Adelie, Biscoe, 37.7, 18.7, 180, 3600, MALE
Adelie, Biscoe, 35.9, 19.2, 189, 3800, FEMALE
Adelie, Biscoe, 38.2, 18.1, 185, 3950, MALE
Adelie, Biscoe, 38.8, 17.2, 180, 3800, MALE
Adelie, Biscoe, 35.3, 18.9, 187, 3800, FEMALE
Adelie, Biscoe, 40.6, 18.6, 183, 3550, MALE
Adelie, Biscoe, 40.5, 17.9, 187, 3200, FEMALE
Adelie, Biscoe, 37.9, 18.6, 172, 3150, FEMALE
```

Empty entries indicate
missing values

Separate columns
with “,” *delimiter*

One observation
per line

CSV Files

Simple, universal

Column labels in first row
(*header*)

Load as Pandas DataFrame

```
species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex
```

```
Adelie, Torgersen, 39.1, 18.7, 181, 3750, MALE
Adelie, Torgersen, 39.5, 17.4, 186, 3800, FEMALE
Adelie, Torgersen, 40.3, 18, 195, 3250, FEMALE
Adelie, Torgersen, , , , ,
Adelie, Torgersen, 36.7, 19.3, 193, 3450, FEMALE
Adelie, Torgersen, 39.3, 20.6, 190, 3650, MALE
Adelie, Torgersen, 38.9, 17.8, 181, 3625, FEMALE
Adelie, Torgersen, 39.2, 19.6, 195, 4675, MALE
Adelie, Torgersen, 34.1, 18.1, 193, 3475,
Adelie, Torgersen, 42, 20.2, 190, 4250,
Adelie, Torgersen, 37.8, 17.1, 186, 3300,
Adelie, Torgersen, 37.8, 17.3, 180, 3700,
Adelie, Torgersen, 41.1, 17.6, 182, 3200, FEMALE
Adelie, Torgersen, 38.6, 21.2, 191, 3800, MALE
Adelie, Torgersen, 34.6, 21.1, 198, 4400, MALE
Adelie, Torgersen, 36.6, 17.8, 185, 3700, FEMALE
Adelie, Torgersen, 38.7, 19, 195, 3450, FEMALE
Adelie, Torgersen, 42.5, 20.7, 197, 4500, MALE
Adelie, Torgersen, 34.4, 18.4, 184, 3325, FEMALE
Adelie, Torgersen, 46, 21.5, 194, 4200, MALE
Adelie, Biscoe, 37.8, 18.3, 174, 3400, FEMALE
Adelie, Biscoe, 37.7, 18.7, 180, 3600, MALE
Adelie, Biscoe, 35.9, 19.2, 189, 3800, FEMALE
Adelie, Biscoe, 38.2, 18.1, 185, 3950, MALE
Adelie, Biscoe, 38.8, 17.2, 180, 3800, MALE
Adelie, Biscoe, 35.3, 18.9, 187, 3800, FEMALE
Adelie, Biscoe, 40.6, 18.6, 183, 3550, MALE
Adelie, Biscoe, 40.5, 17.9, 187, 3200, FEMALE
Adelie, Biscoe, 37.9, 18.6, 172, 3150, FEMALE
```

Empty entries indicate
missing values

Separate columns
with “,” *delimiter*

One observation
per line

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female
...
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	Female
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	Male
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	Female
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	Male

CSV Files

Simple, universal

Column labels in first row
(*header*)

Load as Pandas DataFrame

```
species,island,bill_length_mm,bill_depth_mm,flipper_length_mm,body_mass_g,sex
Adelie,Torgersen,39.1,18.7,181,3750,MALE
Adelie,Torgersen,39.5,17.4,186,3800,FEMALE
Adelie,Torgersen,40.3,18,195,3250,FEMALE
Adelie,Torgersen,,,,,
Adelie,Torgersen,36.7,19.3,193,3450,FEMALE
Adelie,Torgersen,39.3,20.6,190,3650,MALE
Adelie,Torgersen,38.9,17.8,181,3625,FEMALE
Adelie,Torgersen,39.2,19.6,195,4675,MALE
Adelie,Torgersen,34.1,18.1,193,3475,
Adelie,Torgersen,42,20.2,190,4250,
Adelie,Torgersen,37.8,17.1,186,3300,
Adelie,Torgersen,37.8,17.3,180,3700,
Adelie,Torgersen,41.1,17.6,182,3200,FEMALE
Adelie,Torgersen,38.6,21.2,191,3800,MALE
Adelie,Torgersen,34.6,21.1,198,4400,MALE
Adelie,Torgersen,36.6,17.8,185,3700,FEMALE
Adelie,Torgersen,38.7,19,195,3450,FEMALE
Adelie,Torgersen,42.5,20.7,197,4500,MALE
Adelie,Torgersen,34.4,18.4,184,3325,FEMALE
Adelie,Torgersen,46,21.5,194,4200,MALE
Adelie,Biscoe,37.8,18.3,174,3400,FEMALE
Adelie,Biscoe,37.7,18.7,180,3600,MALE
Adelie,Biscoe,35.9,19.2,189,3800,FEMALE
Adelie,Biscoe,38.2,18.1,185,3950,MALE
Adelie,Biscoe,38.8,17.2,180,3800,MALE
Adelie,Biscoe,35.3,18.9,187,3800,FEMALE
Adelie,Biscoe,40.6,18.6,183,3550,MALE
Adelie,Biscoe,40.5,17.9,187,3200,FEMALE
Adelie,Biscoe,37.9,18.6,172,3150,FEMALE
```

Empty entries indicate
missing values

Separate columns
with “,” *delimiter*

One observation
per line

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female
...
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	Female
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	Male
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	Female
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	Male

Pandas missing
values as *NaN*

CSV Files

Simple, universal

Column labels in first row
(*header*)

Load as Pandas DataFrame

```
species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex
Adelie, Torgersen, 39.1, 18.7, 181, 3750, MALE
Adelie, Torgersen, 39.5, 17.4, 186, 3800, FEMALE
Adelie, Torgersen, 40.3, 18, 195, 3250, FEMALE
Adelie, Torgersen, , , , ,
Adelie, Torgersen, 36.7, 19.3, 193, 3450, FEMALE
Adelie, Torgersen, 39.3, 20.6, 190, 3650, MALE
Adelie, Torgersen, 38.9, 17.8, 181, 3625, FEMALE
Adelie, Torgersen, 39.2, 19.6, 195, 4675, MALE
Adelie, Torgersen, 34.1, 18.1, 193, 3475,
Adelie, Torgersen, 42, 20.2, 190, 4250,
Adelie, Torgersen, 37.8, 17.1, 186, 3300,
Adelie, Torgersen, 37.8, 17.3, 180, 3700,
Adelie, Torgersen, 41.1, 17.6, 182, 3200, FEMALE
Adelie, Torgersen, 38.6, 21.2, 191, 3800, MALE
Adelie, Torgersen, 34.6, 21.1, 198, 4400, MALE
Adelie, Torgersen, 36.6, 17.8, 185, 3700, FEMALE
Adelie, Torgersen, 38.7, 19, 195, 3450, FEMALE
Adelie, Torgersen, 42.5, 20.7, 197, 4500, MALE
Adelie, Torgersen, 34.4, 18.4, 184, 3325, FEMALE
Adelie, Torgersen, 46, 21.5, 194, 4200, MALE
Adelie, Biscoe, 37.8, 18.3, 174, 3400, FEMALE
Adelie, Biscoe, 37.7, 18.7, 180, 3600, MALE
Adelie, Biscoe, 35.9, 19.2, 189, 3800, FEMALE
Adelie, Biscoe, 38.2, 18.1, 185, 3950, MALE
Adelie, Biscoe, 38.8, 17.2, 180, 3800, MALE
Adelie, Biscoe, 35.3, 18.9, 187, 3800, FEMALE
Adelie, Biscoe, 40.6, 18.6, 183, 3550, MALE
Adelie, Biscoe, 40.5, 17.9, 187, 3200, FEMALE
Adelie, Biscoe, 37.9, 18.6, 172, 3150, FEMALE
```

Empty entries indicate
missing values

Separate columns
with “,” *delimiter*

One observation
per line

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female
...
339	Gentoo	Biscoe	NaN	NaN	NaN	NaN	NaN
340	Gentoo	Biscoe	46.8	14.3	215.0	4850.0	Female
341	Gentoo	Biscoe	50.4	15.7	222.0	5750.0	Male
342	Gentoo	Biscoe	45.2	14.8	212.0	5200.0	Female
343	Gentoo	Biscoe	49.9	16.1	213.0	5400.0	Male

```
RangeIndex: 344 entries, 0 to 343
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   species                344 non-null    object
1   island                 344 non-null    object
2   bill_length_mm        342 non-null    float64
3   bill_depth_mm         342 non-null    float64
4   flipper_length_mm     342 non-null    float64
5   body_mass_g           342 non-null    float64
6   sex                    333 non-null    object
dtypes: float64(4), object(3)
memory usage: 18.9+ KB
```

Pandas missing
values as *NaN*

Column types *inferred* from data

CSV Files

Not at all simple, definitely not universal

```
Imports,Exports,Balance
"20,527", "23,866", "3,340"
"21,294", "23,039", "1,745"
"22,839", "22,978", 139
"23,912", "24,096", 184
"25,669", "24,142", "-1,527"
"26,413", "25,176", "-1,237"
"26,982", "26,006", -976
"28,197", "25,726", "-2,471"
"28,342", "25,503", "-2,840"
"26,706", "25,454", "-1,252"
"25,814", "24,366", "-1,448"
"24,243", "22,058", "-2,185"
"24,785", "22,570", "-2,215"
"26,211", "22,205", "-4,005"
"27,645", "22,157", "-5,488"
"29,260", "22,172", "-7,088"
```

What if our values
have commas?

CSV Files

Not at all simple, definitely not universal

```
Imports,Exports,Balance
"20,527", "23,866", "3,340"
"21,294", "23,039", "1,745"
"22,839", "22,978", 139
"23,912", "24,096", 184
"25,669", "24,142", "-1,527"
"26,413", "25,176", "-1,237"
"26,982", "26,006", -976
"28,197", "25,726", "-2,471"
"28,342", "25,503", "-2,840"
"26,706", "25,454", "-1,252"
"25,814", "24,366", "-1,448"
"24,243", "22,058", "-2,185"
"24,785", "22,570", "-2,215"
"26,211", "22,205", "-4,005"
"27,645", "22,157", "-5,488"
"29,260", "22,172", "-7,088"
```

What if our values
have commas?

Need to add quotes (“”)

What if our values have quotes?

CSV Files

Not at all simple, definitely not universal

```
Imports,Exports,Balance
"20,527", "23,866", "3,340"
"21,294", "23,039", "1,745"
"22,839", "22,978", 139
"23,912", "24,096", 184
"25,669", "24,142", "-1,527"
"26,413", "25,176", "-1,237"
"26,982", "26,006", -976
"28,197", "25,726", "-2,471"
"28,342", "25,503", "-2,840"
"26,706", "25,454", "-1,252"
"25,814", "24,366", "-1,448"
"24,243", "22,058", "-2,185"
"24,785", "22,570", "-2,215"
"26,211", "22,205", "-4,005"
"27,645", "22,157", "-5,488"
"29,260", "22,172", "-7,088"
```

What if our values
have commas?

Need to add quotes (“”)
What if our values have quotes?

Are these strings or
numbers?

CSV Files

Not at all simple, definitely not universal

```
species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex
Adelie, Torgersen, 39.1, 18.7, 181, 3750, MALE
Adelie, Torgersen, 39.5, 17.4, 186, 3800, FEMALE
Adelie, Torgersen, 40.3, 18, 195, 3250, FEMALE
Adelie, Torgersen, NaN, NaN, NaN, NaN,
Adelie, Torgersen, 36.7, 19.3, 193, 3450, FEMALE
Adelie, Torgersen, 39.3, 20.6, 190, 3650, MALE
```

Strings or missing values?

Header may not always exist!

```
species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex
Adelie, Torgersen, 39.1, 18.7, 181, 3750, MALE
Adelie, Torgersen, 39.5, 17.4, 186, 3800, FEMALE
Adelie, Torgersen, 40.3, 18, 195, 3250, FEMALE
Adelie, Torgersen, , , , ,
Adelie, Torgersen, 36.7, 19.3, 193, 3450, FEMALE
Adelie, Torgersen, 39.3, 20.6, 190, 3650, MALE
```

Strip leading whitespace?

CSV Files

Dates & times are a mess!

1	date	pressure	temperature	wind
2	2010-01-01T01:00:00	1016.6	4.0	3.8
3	2010-01-01T02:00:00	1016.6	3.9	3.8
4	2010-01-01T03:00:00	1016.7	3.8	3.8
5	2010-01-01T04:00:00	1016.7	3.8	3.7
6	2010-01-01T05:00:00	1016.5	3.7	3.8
7	2010-01-01T06:00:00	1016.4	3.7	3.8
8	2010-01-01T07:00:00	1016.5	3.7	3.9
9	2010-01-01T08:00:00	1016.6	3.7	3.9
10	2010-01-01T09:00:00	1017.0	4.0	3.9
11	2010-01-01T10:00:00	1017.4	4.5	3.9
12	2010-01-01T11:00:00	1017.4	5.2	3.9
13	2010-01-01T12:00:00	1017.1	5.8	4.0
14	2010-01-01T13:00:00	1016.7	6.2	4.1

1	symbol	date	price
2	MSFT	Jan 1 2000	39.81
3	MSFT	Feb 1 2000	36.35
4	MSFT	Mar 1 2000	43.22
5	MSFT	Apr 1 2000	28.37
6	MSFT	May 1 2000	25.45
7	MSFT	Jun 1 2000	32.54
8	MSFT	Jul 1 2000	28.4
9	MSFT	Aug 1 2000	28.4
10	MSFT	Sep 1 2000	24.53
11	MSFT	Oct 1 2000	28.02
12	MSFT	Nov 1 2000	23.34
13	MSFT	Dec 1 2000	17.65
14	MSFT	Jan 1 2001	24.84

1	time	count
2	2015/01/01 01:00:00	2
3	2015/01/01 04:00:00	3
4	2015/01/01 05:00:00	1
5	2015/01/01 08:00:00	1
6	2015/01/01 09:00:00	3
7	2015/01/01 11:00:00	1
8	2015/01/01 14:00:00	3
9	2015/01/01 19:00:00	1
10	2015/01/01 22:00:00	1
11	2015/01/02 01:00:00	2
12	2015/01/02 09:00:00	1
13	2015/01/02 10:00:00	3
14	2015/01/02 12:00:00	2

1	UN member since	World bank region	World bank, 4 income groups 2017
2	1/11/1945	East Asia & Pacific	High income
3	21/9/1984	East Asia & Pacific	High income
4	14/12/1955	East Asia & Pacific	Lower middle income
5	24/10/1945	East Asia & Pacific	Upper middle income
6	13/10/1970	East Asia & Pacific	Upper middle income
7		East Asia & Pacific	High income
8	28/9/1950	East Asia & Pacific	Lower middle income
9	18/12/1956	East Asia & Pacific	High income
10	14/9/1999	East Asia & Pacific	Lower middle income
11	17/9/1991	East Asia & Pacific	Low income
11	17/9/1991	East Asia & Pacific	High income
12	14/12/1955	East Asia & Pacific	Lower middle income
13	17/9/1957	East Asia & Pacific	Upper middle income
14	17/9/1991	East Asia & Pacific	Upper middle income

ISO Date format
rarely used!

CSV Files

The pandas approach

```
pandas.read_csv(filepath_or_buffer, *, sep=<no_default>,
delimiter=None, header='infer', names=<no_default>,
index_col=None, usecols=None, dtype=None, engine=None,
converters=None, true_values=None, false_values=None,
skipinitialspace=False, skiprows=None, skipfooter=0,
nrows=None, na_values=None, keep_default_na=True,
na_filter=True, verbose=<no_default>, skip_blank_lines=True,
parse_dates=None, infer_datetime_format=<no_default>,
keep_date_col=<no_default>, date_parser=<no_default>,
date_format=None, dayfirst=False, cache_dates=True,
iterator=False, chunksize=None, compression='infer',
thousands=None, decimal='.', lineterminator=None,
quotechar='"', quoting=0, doublequote=True, escapechar=None,
comment=None, encoding=None, encoding_errors='strict',
dialect=None, on_bad_lines='error',
delim_whitespace=<no_default>, low_memory=True,
memory_map=False, float_precision=None, storage_options=None,
dtype_backend=<no_default>)
```

[\[source\]](#)

TSV Files

Alternative delimiters sometime used (e.g. TAB)

```
species,island,bill_length_mm,bill_depth_mm,flipper_length_mm,body_mass_g,sex
Adelie,Torgersen,39.1,18.7,181,3750,MALE
Adelie,Torgersen,39.5,17.4,186,3800,FEMALE
Adelie,Torgersen,40.3,18,195,3250,FEMALE
Adelie,Torgersen,,,,
Adelie,Torgersen,36.7,19.3,193,3450,FEMALE
Adelie,Torgersen,39.3,20.6,190,3650,MALE
Adelie,Torgersen,38.9,17.8,181,3625,FEMALE
Adelie,Torgersen,39.2,19.6,195,4675,MALE
Adelie,Torgersen,34.1,18.1,193,3475,
Adelie,Torgersen,42,20.2,190,4250,
Adelie,Torgersen,37.8,17.1,186,3300,
Adelie,Torgersen,37.8,17.3,180,3700,
Adelie,Torgersen,41.1,17.6,182,3200,FEMALE
Adelie,Torgersen,38.6,21.2,191,3800,MALE
Adelie,Torgersen,34.6,21.1,198,4400,MALE
Adelie,Torgersen,36.6,17.8,185,3700,FEMALE
Adelie,Torgersen,38.7,19,195,3450,FEMALE
Adelie,Torgersen,42.5,20.7,197,4500,MALE
Adelie,Torgersen,34.4,18.4,184,3325,FEMALE
Adelie,Torgersen,46,21.5,194,4200,MALE
Adelie,Biscoe,37.8,18.3,174,3400,FEMALE
Adelie,Biscoe,37.7,18.7,180,3600,MALE
Adelie,Biscoe,35.9,19.2,189,3800,FEMALE
Adelie,Biscoe,38.2,18.1,185,3950,MALE
Adelie,Biscoe,37.7,18.7,180,3600,MALE
Adelie,Biscoe,35.9,19.2,189,3800,FEMALE
Adelie,Biscoe,38.2,18.1,185,3950,MALE
Adelie,Biscoe,38.8,17.2,180,3800,MALE
Adelie,Biscoe,35.3,18.9,187,3800,FEMALE
Adelie,Biscoe,40.6,18.6,183,3550,MALE
Adelie,Biscoe,40.5,17.9,187,3200,FEMALE
Adelie,Biscoe,37.9,18.6,172,3150,FEMALE
```

CSV File

```
species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g sex
Adelie Torgersen 39.1 18.7 181 3750 MALE
Adelie Torgersen 39.5 17.4 186 3800 FEMALE
Adelie Torgersen 40.3 18 195 3250 FEMALE
Adelie Torgersen
Adelie Torgersen 36.7 19.3 193 3450 FEMALE
Adelie Torgersen 39.3 20.6 190 3650 MALE
Adelie Torgersen 38.9 17.8 181 3625 FEMALE
Adelie Torgersen 39.2 19.6 195 4675 MALE
Adelie Torgersen 34.1 18.1 193 3475
Adelie Torgersen 42 20.2 190 4250
Adelie Torgersen 37.8 17.1 186 3300
Adelie Torgersen 37.8 17.3 180 3700
Adelie Torgersen 41.1 17.6 182 3200 FEMALE
Adelie Torgersen 38.6 21.2 191 3800 MALE
Adelie Torgersen 34.6 21.1 198 4400 MALE
Adelie Torgersen 36.6 17.8 185 3700 FEMALE
Adelie Torgersen 38.7 19 195 3450 FEMALE
Adelie Torgersen 42.5 20.7 197 4500 MALE
Adelie Torgersen 34.4 18.4 184 3325 FEMALE
Adelie Torgersen 46 21.5 194 4200 MALE
Adelie Biscoe 37.8 18.3 174 3400 FEMALE
Adelie Biscoe 37.7 18.7 180 3600 MALE
Adelie Biscoe 35.9 19.2 189 3800 FEMALE
Adelie Biscoe 38.2 18.1 185 3950 MALE
Adelie Biscoe 37.7 18.7 180 3600 MALE
Adelie Biscoe 35.9 19.2 189 3800 FEMALE
Adelie Biscoe 40.6 18.6 183 3550 MALE
Adelie Biscoe 40.5 17.9 187 3200 FEMALE
Adelie Biscoe 37.9 18.6 172 3150 FEMALE
Adelie Biscoe 40.5 18.9 180 3950 MALE
Adelie Dream 39.5 16.7 178 3250 FEMALE
Adelie Dream 37.2 18.1 178 3900 MALE
Adelie Dream 39.5 17.8 188 3300 FEMALE
```

TSV File

JavaScript Object Notation (JSON)

Common CSV alternative

```
species,island,bill_length_mm,bill_depth_mm,flipper_length_mm,body_mass_g,sex
Adelie,Torgersen,39.1,18.7,181,3750,MALE
Adelie,Torgersen,39.5,17.4,186,3800,FEMALE
Adelie,Torgersen,40.3,18,195,3250,FEMALE
Adelie,Torgersen,,,,,
Adelie,Torgersen,36.7,19.3,193,3450,FEMALE
Adelie,Torgersen,39.3,20.6,190,3650,MALE
Adelie,Torgersen,38.9,17.8,181,3625,FEMALE
Adelie,Torgersen,39.2,19.6,195,4675,MALE
Adelie,Torgersen,34.1,18.1,193,3475,
Adelie,Torgersen,42,20.2,190,4250,
Adelie,Torgersen,37.8,17.1,186,3300,
Adelie,Torgersen,37.8,17.3,180,3700,
Adelie,Torgersen,41.1,17.6,182,3200,FEMALE
Adelie,Torgersen,38.6,21.2,191,3800,MALE
Adelie,Torgersen,34.6,21.1,198,4400,MALE
Adelie,Torgersen,36.6,17.8,185,3700,FEMALE
Adelie,Torgersen,38.7,19,195,3450,FEMALE
Adelie,Torgersen,42.5,20.7,197,4500,MALE
Adelie,Torgersen,34.4,18.4,184,3325,FEMALE
Adelie,Torgersen,46,21.5,194,4200,MALE
Adelie,Biscoe,37.8,18.3,174,3400,FEMALE
Adelie,Biscoe,37.7,18.7,180,3600,MALE
Adelie,Biscoe,35.9,19.2,189,3800,FEMALE
Adelie,Biscoe,38.2,18.1,185,3950,MALE
Adelie,Biscoe,38.8,17.2,180,3800,MALE
Adelie,Biscoe,35.3,18.9,187,3800,FEMALE
Adelie,Biscoe,40.6,18.6,183,3550,MALE
Adelie,Biscoe,40.5,17.9,187,3200,FEMALE
Adelie,Biscoe,37.9,18.6,172,3150,FEMALE
```

CSV File

```
[
  {
    "Species": "Adelie",
    "Island": "Torgersen",
    "Beak Length (mm)": 39.1,
    "Beak Depth (mm)": 18.7,
    "Flipper Length (mm)": 181,
    "Body Mass (g)": 3750,
    "Sex": "MALE"
  },
  {
    "Species": "Adelie",
    "Island": "Torgersen",
    "Beak Length (mm)": 39.5,
    "Beak Depth (mm)": 17.4,
    "Flipper Length (mm)": 186,
    "Body Mass (g)": 3800,
    "Sex": "FEMALE"
  },
  {
    "Species": "Adelie",
    "Island": "Torgersen",
    "Beak Length (mm)": 40.3,
    "Beak Depth (mm)": 18,
    "Flipper Length (mm)": 195,
    "Body Mass (g)": 3250,
    "Sex": "FEMALE"
  },
  {
    "Species": "Adelie",
    "Island": "Torgersen",
    "Beak Length (mm)": null,
    "Beak Depth (mm)": null,
    "Flipper Length (mm)": null,
    "Body Mass (g)": null,
    "Sex": null
  },
],
```

JSON File

Nested lists and dictionaries (*arrays* and *objects* in JS)

Records format: List of dicts mapping column name to values

Not universal! JSON can have other structures

Other common formats

Many can be loaded with Pandas!



Excel Files



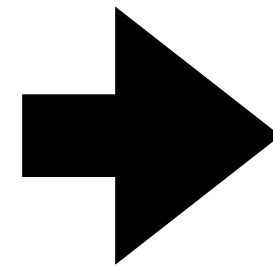
SQL Databases

Table operations

Filtering data

species == "Gentoo"

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	45.2	15.8	215.0	5300.0	Male
Adelie	Dream	40.6	17.2	187.0	3475.0	Male
Adelie	Torgersen	34.1	18.1	193.0	3475.0	None
Adelie	Torgersen	37.7	19.8	198.0	3500.0	Male
Gentoo	Biscoe	50.8	17.3	228.0	5600.0	Male
Adelie	Biscoe	38.2	20.0	190.0	3900.0	Male
Adelie	Biscoe	45.6	20.3	191.0	4600.0	Male
Gentoo	Biscoe	40.9	13.7	214.0	4650.0	Female
Gentoo	Biscoe	44.0	13.6	208.0	4350.0	Female
Chinstrap	Dream	51.3	18.2	197.0	3750.0	Male
Adelie	Biscoe	41.3	21.1	195.0	4400.0	Male
Adelie	Torgersen	33.5	19.0	190.0	3600.0	Female
Gentoo	Biscoe	46.2	14.5	209.0	4800.0	Female
Adelie	Dream	35.7	18.0	202.0	3550.0	Female
Adelie	Torgersen	40.9	16.8	191.0	3700.0	Female
Chinstrap	Dream	50.2	18.8	202.0	3800.0	Male
Adelie	Dream	41.1	17.5	190.0	3900.0	Male

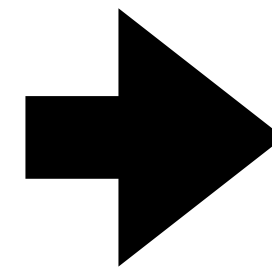


species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	45.2	15.8	215.0	5300.0	Male
Gentoo	Biscoe	50.8	17.3	228.0	5600.0	Male
Gentoo	Biscoe	40.9	13.7	214.0	4650.0	Female
Gentoo	Biscoe	44.0	13.6	208.0	4350.0	Female
Gentoo	Biscoe	46.2	14.5	209.0	4800.0	Female

Filtering data

species == “Adelie” & body_mass_g > 3600

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	45.2	15.8	215.0	5300.0	Male
Adelie	Dream	40.6	17.2	187.0	3475.0	Male
Adelie	Torgersen	34.1	18.1	193.0	3475.0	None
Adelie	Torgersen	37.7	19.8	198.0	3500.0	Male
Gentoo	Biscoe	50.8	17.3	228.0	5600.0	Male
Adelie	Biscoe	38.2	20.0	190.0	3900.0	Male
Adelie	Biscoe	45.6	20.3	191.0	4600.0	Male
Gentoo	Biscoe	40.9	13.7	214.0	4650.0	Female
Gentoo	Biscoe	44.0	13.6	208.0	4350.0	Female
Chinstrap	Dream	51.3	18.2	197.0	3750.0	Male
Adelie	Biscoe	41.3	21.1	195.0	4400.0	Male
Adelie	Torgersen	33.5	19.0	190.0	3600.0	Female
Gentoo	Biscoe	46.2	14.5	209.0	4800.0	Female
Adelie	Dream	35.7	18.0	202.0	3550.0	Female
Adelie	Torgersen	40.9	16.8	191.0	3700.0	Female
Chinstrap	Dream	50.2	18.8	202.0	3800.0	Male

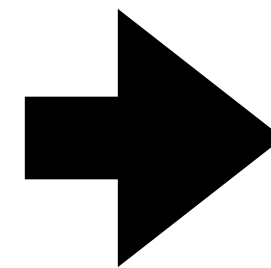


species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Adelie	Biscoe	38.2	20.0	190.0	3900.0	Male
Adelie	Biscoe	45.6	20.3	191.0	4600.0	Male
Adelie	Biscoe	41.3	21.1	195.0	4400.0	Male
Adelie	Torgersen	40.9	16.8	191.0	3700.0	Female
Adelie	Dream	41.1	17.5	190.0	3900.0	Male

Sorting data

By body_mass_g

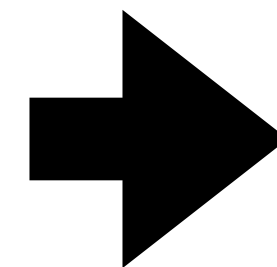
species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	45.2	15.8	215.0	5300.0	Male
Adelie	Dream	40.6	17.2	187.0	3475.0	Male
Adelie	Torgersen	34.1	18.1	193.0	3475.0	None
Adelie	Torgersen	37.7	19.8	198.0	3500.0	Male
Gentoo	Biscoe	50.8	17.3	228.0	5600.0	Male
Adelie	Biscoe	38.2	20.0	190.0	3900.0	Male
Adelie	Biscoe	45.6	20.3	191.0	4600.0	Male
Gentoo	Biscoe	40.9	13.7	214.0	4650.0	Female
Gentoo	Biscoe	44.0	13.6	208.0	4350.0	Female
Chinstrap	Dream	51.3	18.2	197.0	3750.0	Male
Adelie	Biscoe	41.3	21.1	195.0	4400.0	Male
Adelie	Torgersen	33.5	19.0	190.0	3600.0	Female
Gentoo	Biscoe	46.2	14.5	209.0	4800.0	Female
Adelie	Dream	35.7	18.0	202.0	3550.0	Female
Adelie	Torgersen	40.9	16.8	191.0	3700.0	Female
Chinstrap	Dream	50.2	18.8	202.0	3800.0	Male
Adelie	Dream	41.1	17.5	190.0	3900.0	Male



species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	50.8	17.3	228.0	5600.0	Male
Gentoo	Biscoe	45.2	15.8	215.0	5300.0	Male
Gentoo	Biscoe	46.2	14.5	209.0	4800.0	Female
Gentoo	Biscoe	40.9	13.7	214.0	4650.0	Female
Adelie	Biscoe	45.6	20.3	191.0	4600.0	Male
Adelie	Biscoe	41.3	21.1	195.0	4400.0	Male
Gentoo	Biscoe	44.0	13.6	208.0	4350.0	Female
Adelie	Biscoe	38.2	20.0	190.0	3900.0	Male
Adelie	Dream	41.1	17.5	190.0	3900.0	Male
Chinstrap	Dream	50.2	18.8	202.0	3800.0	Male
Chinstrap	Dream	51.3	18.2	197.0	3750.0	Male
Adelie	Torgersen	40.9	16.8	191.0	3700.0	Female
Adelie	Torgersen	33.5	19.0	190.0	3600.0	Female
Adelie	Dream	35.7	18.0	202.0	3550.0	Female
Adelie	Torgersen	37.7	19.8	198.0	3500.0	Male
Adelie	Dream	40.6	17.2	187.0	3475.0	Male
Adelie	Torgersen	34.1	18.1	193.0	3475.0	None

Selecting variables

species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
Gentoo	Biscoe	45.2	15.8	215.0	5300.0	Male
Adelie	Dream	40.6	17.2	187.0	3475.0	Male
Adelie	Torgersen	34.1	18.1	193.0	3475.0	None
Adelie	Torgersen	37.7	19.8	198.0	3500.0	Male
Gentoo	Biscoe	50.8	17.3	228.0	5600.0	Male
Adelie	Biscoe	38.2	20.0	190.0	3900.0	Male
Adelie	Biscoe	45.6	20.3	191.0	4600.0	Male
Gentoo	Biscoe	40.9	13.7	214.0	4650.0	Female
Gentoo	Biscoe	44.0	13.6	208.0	4350.0	Female
Chinstrap	Dream	51.3	18.2	197.0	3750.0	Male
Adelie	Biscoe	41.3	21.1	195.0	4400.0	Male
Adelie	Torgersen	33.5	19.0	190.0	3600.0	Female
Gentoo	Biscoe	46.2	14.5	209.0	4800.0	Female
Adelie	Dream	35.7	18.0	202.0	3550.0	Female
Adelie	Torgersen	40.9	16.8	191.0	3700.0	Female
Chinstrap	Dream	50.2	18.8	202.0	3800.0	Male
Adelie	Dream	41.1	17.5	190.0	3900.0	Male



species	island	body_mass_g
Gentoo	Biscoe	5300.0
Adelie	Dream	3475.0
Adelie	Torgersen	3475.0
Adelie	Torgersen	3500.0
Gentoo	Biscoe	5600.0
Adelie	Biscoe	3900.0
Adelie	Biscoe	4600.0
Gentoo	Biscoe	4650.0
Gentoo	Biscoe	4350.0
Chinstrap	Dream	3750.0
Adelie	Biscoe	4400.0
Adelie	Torgersen	3600.0
Gentoo	Biscoe	4800.0
Adelie	Dream	3550.0
Adelie	Torgersen	3700.0
Chinstrap	Dream	3800.0

Pivoting data

Id = "origin", values = "count", variable = "destination"

origin	destination	count
ATL	BOS	5990
ATL	JFK	1915
ATL	LAS	4766
ATL	LAX	5406
ATL	ORD	7677
ATL	PHX	4603
ATL	SAN	2857
ATL	SFO	3462
BOS	ATL	5982
BOS	JFK	7372
BOS	LAS	1383
BOS	LAX	2234
BOS	ORD	7085
BOS	PHX	1069
BOS	SAN	608

Pivot



Melt / unpivot

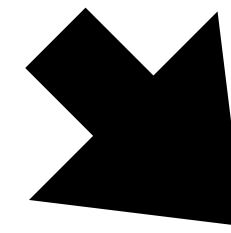
origin	ATL	BOS	JFK	LAS	LAX	ORD	PHX	SAN	SFO
ATL	None	5990.0	1915.0	4766.0	5406.0	7677.0	4603.0	2857.0	3462.0
BOS	5982.0	None	7372.0	1383.0	2234.0	7085.0	1069.0	608.0	2668.0
JFK	2279.0	7513.0	None	4471.0	8078.0	3672.0	2254.0	1630.0	6591.0
LAS	4761.0	1383.0	4481.0	None	11729.0	5235.0	10626.0	5895.0	6994.0
LAX	5402.0	2441.0	8058.0	11773.0	None	7333.0	9897.0	11257.0	13390.0
ORD	7449.0	7133.0	3684.0	5257.0	6974.0	None	5124.0	3308.0	5524.0
PHX	4607.0	1067.0	2256.0	10337.0	9992.0	5130.0	None	7609.0	4884.0
SAN	2855.0	610.0	1630.0	5902.0	11224.0	2948.0	7581.0	None	6769.0
SFO	3098.0	2594.0	6971.0	7007.0	13788.0	5746.0	4870.0	6312.0	None

Inner join On “origin”

origin	city	state
ATL	Atlanta	GA
DFW	Dallas-Fort Worth	TX
LAS	Las Vegas	NV
LAX	Los Angeles	CA
LGA	New York	NY
MIA	Miami	FL
ORD	Chicago	IL

origin	destination	count
LAX	BOS	2441
JFK	SFO	6591
BOS	LAX	2234
SFO	PHX	4870
PHX	ATL	4607
LAX	PHX	9897
LAS	ATL	4761
SFO	ATL	3098
LAS	LAX	11729
ORD	SAN	3308

Keep matches



Keep matches

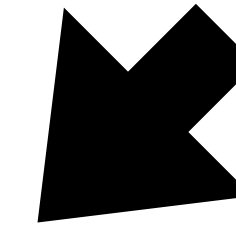
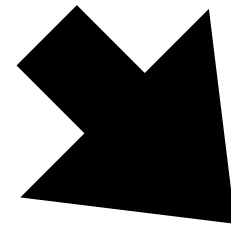
origin	city	state	destination	count
LAS	Las Vegas	NV	ATL	4761
LAS	Las Vegas	NV	LAX	11729
LAX	Los Angeles	CA	BOS	2441
LAX	Los Angeles	CA	PHX	9897
ORD	Chicago	IL	SAN	3308
SFO	San Francisco	CA	PHX	4870
SFO	San Francisco	CA	ATL	3098

Left join On “origin”

origin	city	state
ATL	Atlanta	GA
DFW	Dallas-Fort Worth	TX
LAS	Las Vegas	NV
LAX	Los Angeles	CA
LGA	New York	NY
MIA	Miami	FL
ORD	Chicago	IL
SFO	San Francisco	CA

origin	destination	count
LAX	BOS	2441
JFK	SFO	6591
BOS	LAX	2234
SFO	PHX	4870
PHX	ATL	4607
LAX	PHX	9897
LAS	ATL	4761
SFO	ATL	3098
LAS	LAX	11729
ORD	SAN	3308

Keep all



Keep matches

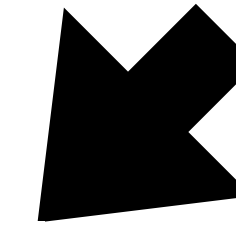
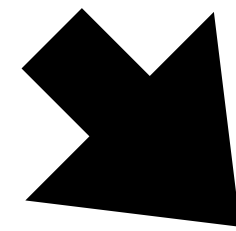
origin	city	state	destination	count
ATL	Atlanta	GA		
DFW	Dallas-Fort Worth	TX		
LAS	Las Vegas	NV	ATL	4761.0
LAS	Las Vegas	NV	LAX	11729.0
LAX	Los Angeles	CA	BOS	2441.0
LAX	Los Angeles	CA	PHX	9897.0
LGA	New York	NY		
MIA	Miami	FL		
ORD	Chicago	IL	SAN	3308.0
SFO	San Francisco	CA	PHX	4870.0
SFO	San Francisco	CA	ATL	3098.0

Outer join On “origin”

origin	city	state
ATL	Atlanta	GA
DFW	Dallas-Fort Worth	TX
LAS	Las Vegas	NV
LAX	Los Angeles	CA
LGA	New York	NY
MIA	Miami	FL
ORD	Chicago	IL

origin	destination	count
LAX	BOS	2441
JFK	SFO	6591
BOS	LAX	2234
SFO	PHX	4870
PHX	ATL	4607
LAX	PHX	9897
LAS	ATL	4761
SFO	ATL	3098
LAS	LAX	11729

Keep all



Keep all

Duplicate if needed!

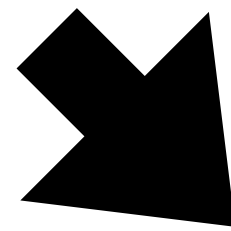
origin	city	state	destination	count
ATL	Atlanta	GA		
BOS			LAX	2234.0
DFW	Dallas-Fort Worth	TX		
JFK			SFO	6591.0
LAS	Las Vegas	NV	ATL	4761.0
LAS	Las Vegas	NV	LAX	11729.0
LAX	Los Angeles	CA	BOS	2441.0
LAX	Los Angeles	CA	PHX	9897.0
LGA	New York	NY		
MIA	Miami	FL		
ORD	Chicago	IL	SAN	3308.0
PHX			ATL	4607.0
SFO	San Francisco	CA	PHX	4870.0
SFO	San Francisco	CA	ATL	3098.0

Right join On “origin”

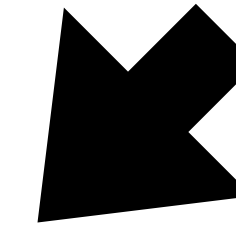
origin	city	state
ATL	Atlanta	GA
DFW	Dallas-Fort Worth	TX
LAS	Las Vegas	NV
LAX	Los Angeles	CA
LGA	New York	NY
MIA	Miami	FL
ORD	Chicago	IL
SFO	San Francisco	CA

origin	destination	count
LAX	BOS	2441
JFK	SFO	6591
BOS	LAX	2234
SFO	PHX	4870
PHX	ATL	4607
LAX	PHX	9897
LAS	ATL	4761
SFO	ATL	3098
LAS	LAX	11729

Keep matches



Keep all

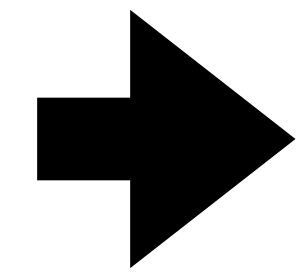


Duplicate if needed!

origin	city	state	destination	count
LAX	Los Angeles	CA	BOS	2441
JFK			SFO	6591
BOS			LAX	2234
SFO	San Francisco	CA	PHX	4870
PHX			ATL	4607
LAX	Los Angeles	CA	PHX	9897
LAS	Las Vegas	NV	ATL	4761
SFO	San Francisco	CA	ATL	3098
LAS	Las Vegas	NV	LAX	11729

Grouping By “species”

species	flipper_length_mm	body_mass_g
Gentoo	215.0	5300.0
Adelie	187.0	3475.0
Adelie	193.0	3475.0
Adelie	198.0	3500.0
Gentoo	228.0	5600.0
Adelie	190.0	3900.0
Adelie	191.0	4600.0
Gentoo	214.0	4650.0
Gentoo	208.0	4350.0
Chinstrap	197.0	3750.0
Adelie	195.0	4400.0
Adelie	190.0	3600.0
Gentoo	209.0	4800.0
Adelie	202.0	3550.0
Adelie	191.0	3700.0
Chinstrap	202.0	3800.0



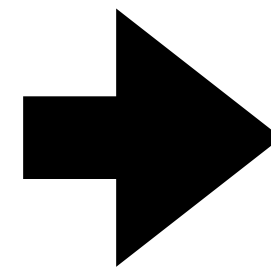
Mean

species	flipper_length_mm	body_mass_g
Gentoo	214.8	4940.0
Adelie	192.7	3810.0
Chinstrap	199.5	3775.0

Grouping

By “species” and island

species	island	flipper_length_mm	body_mass_g
Gentoo	Biscoe	215.0	5300.0
Adelie	Dream	187.0	3475.0
Adelie	Torgersen	193.0	3475.0
Adelie	Torgersen	198.0	3500.0
Gentoo	Biscoe	228.0	5600.0
Adelie	Biscoe	190.0	3900.0
Adelie	Biscoe	191.0	4600.0
Gentoo	Biscoe	214.0	4650.0
Gentoo	Biscoe	208.0	4350.0
Chinstrap	Dream	197.0	3750.0
Adelie	Biscoe	195.0	4400.0
Adelie	Torgersen	190.0	3600.0
Gentoo	Biscoe	209.0	4800.0
Adelie	Dream	202.0	3550.0
Adelie	Torgersen	191.0	3700.0
Chinstrap	Dream	202.0	3800.0
Adelie	Dream	190.0	3900.0



Mean

species	island	flipper_length_mm	body_mass_g
Gentoo	Biscoe	214.8	4940.0
Chinstrap	Dream	199.5	3775.0
Adelie	Dream	193.0	3641.6666666666665
Adelie	Torgersen	193.0	3568.75
Adelie	Biscoe	192.0	4300.0

Deriving

Create `body_mass_kg = body_mass_g / 1000`.

species	flipper_length_mm	body_mass_g
Gentoo	215.0	5300.0
Adelie	187.0	3475.0
Adelie	193.0	3475.0
Adelie	198.0	3500.0
Gentoo	228.0	5600.0
Adelie	190.0	3900.0
Adelie	191.0	4600.0
Gentoo	214.0	4650.0
Gentoo	208.0	4350.0
Chinstrap	197.0	3750.0
Adelie	195.0	4400.0
Adelie	190.0	3600.0
Gentoo	209.0	4800.0
Adelie	202.0	3550.0
Adelie	191.0	3700.0
Chinstrap	202.0	3800.0
Adelie	190.0	3900.0

species	flipper_length_mm	body_mass_g	body_mass_kg
Gentoo	215.0	5300.0	5.3
Adelie	187.0	3475.0	3.475
Adelie	193.0	3475.0	3.475
Adelie	198.0	3500.0	3.5
Gentoo	228.0	5600.0	5.6
Adelie	190.0	3900.0	3.9
Adelie	191.0	4600.0	4.6
Gentoo	214.0	4650.0	4.65
Gentoo	208.0	4350.0	4.35
Chinstrap	197.0	3750.0	3.75
Adelie	195.0	4400.0	4.4
Adelie	190.0	3600.0	3.6
Gentoo	209.0	4800.0	4.8
Adelie	202.0	3550.0	3.55
Adelie	191.0	3700.0	3.7
Chinstrap	202.0	3800.0	3.8