


```
▶ In [2]: import pandas as pd
surveys_df = pd.read_csv("surveys.csv",
                        keep_default_na=False, na_values=[""])
surveys_df
```

Out[2]:

	record_id	month	day	year	plot_id	species_id	sex	hindfoot_length	weight
0	1	7	16	1977	2	NL	M	32.0	NaN
1	2	7	16	1977	3	NL	M	33.0	NaN
2	3	7	16	1977	2	DM	F	37.0	NaN
3	4	7	16	1977	7	DM	M	36.0	NaN
4	5	7	16	1977	3	DM	M	35.0	NaN
5	6	7	16	1977	1	PF	M	14.0	NaN
6	7	7	16	1977	2	PE	F	NaN	NaN
7	8	7	16	1977	1	DM	M	37.0	NaN
8	9	7	16	1977	1	DM	F	34.0	NaN
9	10	7	16	1977	6	PF	F	20.0	NaN
10	11	7	16	1977	5	DS	F	53.0	NaN
11	12	7	16	1977	7	DM	M	38.0	NaN
12	13	7	16	1977	3	DM	M	35.0	NaN
13	14	7	16	1977	8	DM	NaN	NaN	NaN
14	15	7	16	1977	6	DM	F	36.0	NaN
15	16	7	16	1977	4	DM	F	36.0	NaN
16	17	7	16	1977	3	DS	F	48.0	NaN
17	18	7	16	1977	2	PP	M	22.0	NaN
18	19	7	16	1977	4	PF	NaN	NaN	NaN
19	20	7	17	1977	11	DS	F	48.0	NaN
20	21	7	17	1977	14	DM	F	34.0	NaN
21	22	7	17	1977	15	NL	F	31.0	NaN
22	23	7	17	1977	13	DM	M	36.0	NaN
23	24	7	17	1977	13	SH	M	21.0	NaN
24	25	7	17	1977	9	DM	M	35.0	NaN
25	26	7	17	1977	15	DM	M	31.0	NaN
26	27	7	17	1977	15	DM	M	36.0	NaN
27	28	7	17	1977	11	DM	M	38.0	NaN
28	29	7	17	1977	11	PP	M	NaN	NaN
29	30	7	17	1977	10	DS	F	52.0	NaN
...
35519	35520	12	31	2002	9	SF	NaN	24.0	36.0
35520	35521	12	31	2002	9	DM	M	37.0	48.0
35521	35522	12	31	2002	9	DM	F	35.0	45.0
35522	35523	12	31	2002	9	DM	F	36.0	44.0
35523	35524	12	31	2002	9	PB	F	25.0	27.0
35524	35525	12	31	2002	9	OL	M	21.0	26.0
35525	35526	12	31	2002	8	OT	F	20.0	24.0

	record_id	month	day	year	plot_id	species_id	sex	hindfoot_length	weight
35526	35527	12	31	2002	13	DO	F	33.0	43.0
35527	35528	12	31	2002	13	US	NaN	NaN	NaN
35528	35529	12	31	2002	13	PB	F	25.0	25.0
35529	35530	12	31	2002	13	OT	F	20.0	NaN
35530	35531	12	31	2002	13	PB	F	27.0	NaN
35531	35532	12	31	2002	14	DM	F	34.0	43.0
35532	35533	12	31	2002	14	DM	F	36.0	48.0
35533	35534	12	31	2002	14	DM	M	37.0	56.0
35534	35535	12	31	2002	14	DM	M	37.0	53.0
35535	35536	12	31	2002	14	DM	F	35.0	42.0
35536	35537	12	31	2002	14	DM	F	36.0	46.0
35537	35538	12	31	2002	15	PB	F	26.0	31.0
35538	35539	12	31	2002	15	SF	M	26.0	68.0
35539	35540	12	31	2002	15	PB	F	26.0	23.0
35540	35541	12	31	2002	15	PB	F	24.0	31.0
35541	35542	12	31	2002	15	PB	F	26.0	29.0
35542	35543	12	31	2002	15	PB	F	27.0	34.0
35543	35544	12	31	2002	15	US	NaN	NaN	NaN
35544	35545	12	31	2002	15	AH	NaN	NaN	NaN
35545	35546	12	31	2002	15	AH	NaN	NaN	NaN
35546	35547	12	31	2002	10	RM	F	15.0	14.0
35547	35548	12	31	2002	7	DO	M	36.0	51.0
35548	35549	12	31	2002	5	NaN	NaN	NaN	NaN

35549 rows × 9 columns

```
In [6]: species_df = pd.read_csv("species.csv",  
                                keep_default_na=False, na_values=[""])  
species_df
```

Out[6]:

	species_id	genus	species	taxa
0	AB	Amphispiza	bilineata	Bird
1	AH	Ammospermophilus	harrisi	Rodent
2	AS	Ammodramus	savannarum	Bird
3	BA	Baiomys	taylori	Rodent
4	CB	Campylorhynchus	brunneicapillus	Bird
5	CM	Calamospiza	melanocorys	Bird
6	CQ	Callipepla	squamata	Bird
7	CS	Crotalus	scutalatus	Reptile
8	CT	Cnemidophorus	tigris	Reptile
9	CU	Cnemidophorus	uniparens	Reptile
10	CV	Crotalus	viridis	Reptile
11	DM	Dipodomys	merriami	Rodent
12	DO	Dipodomys	ordii	Rodent
13	DS	Dipodomys	spectabilis	Rodent
14	DX	Dipodomys	sp.	Rodent
15	EO	Eumeces	obsoletus	Reptile
16	GS	Gambelia	silus	Reptile
17	NL	Neotoma	albigula	Rodent
18	NX	Neotoma	sp.	Rodent
19	OL	Onychomys	leucogaster	Rodent
20	OT	Onychomys	torridus	Rodent
21	OX	Onychomys	sp.	Rodent
22	PB	Chaetodipus	baileyi	Rodent
23	PC	Pipilo	chlorurus	Bird
24	PE	Peromyscus	eremicus	Rodent
25	PF	Perognathus	flavus	Rodent
26	PG	Pooecetes	gramineus	Bird
27	PH	Perognathus	hispidus	Rodent
28	PI	Chaetodipus	intermedius	Rodent
29	PL	Peromyscus	leucopus	Rodent
30	PM	Peromyscus	maniculatus	Rodent
31	PP	Chaetodipus	penicillatus	Rodent
32	PU	Pipilo	fuscus	Bird
33	PX	Chaetodipus	sp.	Rodent
34	RF	Reithrodontomys	fulvescens	Rodent
35	RM	Reithrodontomys	megalotis	Rodent
36	RO	Reithrodontomys	montanus	Rodent
37	RX	Reithrodontomys	sp.	Rodent

	species_id	genus	species	taxa
38	SA	Sylvilagus	audubonii	Rabbit
39	SB	Spizella	breweri	Bird
40	SC	Sceloporus	clarki	Reptile
41	SF	Sigmodon	fulviventer	Rodent
42	SH	Sigmodon	hispidus	Rodent
43	SO	Sigmodon	ochrognathus	Rodent
44	SS	Spermophilus	spilosoma	Rodent
45	ST	Spermophilus	tereticaudus	Rodent
46	SU	Sceloporus	undulatus	Reptile
47	SX	Sigmodon	sp.	Rodent
48	UL	Lizard	sp.	Reptile
49	UP	Pipilo	sp.	Bird
50	UR	Rodent	sp.	Rodent
51	US	Sparrow	sp.	Bird
52	ZL	Zonotrichia	leucophrys	Bird
53	ZM	Zenaida	macroura	Bird

```
In [7]: # Read in first 10 lines of surveys table
survey_sub = surveys_df.head(10)
```

```
In [8]: # Grab the last 10 rows
survey_sub_last10 = surveys_df.tail(10)
```

```
In [9]: # Reset the index values to the second dataframe appends properly
survey_sub_last10 = survey_sub_last10.reset_index(drop=True)
# drop=True option avoids adding new index column with old index values
```

```
In [10]: # Stack the DataFrames on top of each other
vertical_stack = pd.concat([survey_sub, survey_sub_last10], axis=0)
```

```
In [11]: # Place the DataFrames side by side
horizontal_stack = pd.concat([survey_sub, survey_sub_last10], axis=1)
```

```
In [12]: # Write DataFrame to CSV
vertical_stack.to_csv('out.csv', index=False)
```

```
In [13]: # For kicks read our output back into Python and make sure all looks good
new_output = pd.read_csv('out.csv', keep_default_na=False, na_values=[""])
```

```
In [14]: # Read in first 10 lines of surveys table
survey_sub = surveys_df.head(10)
```

```
In [16]: species_sub = pd.read_csv('species.csv', keep_default_na=False, na_values=[""])
#这里没有找到/speciesSubset.csv文件，用的是species.csv
```

```
In [17]: species_sub.columns
```

```
Out[17]: Index(['species_id', 'genus', 'species', 'taxa'], dtype='object')
```

```
In [18]: survey_sub.columns
```

```
Out[18]: Index(['record_id', 'month', 'day', 'year', 'plot_id', 'species_id', 'sex',
               'hindfoot_length', 'weight'],
               dtype='object')
```

```
In [19]: merged_inner = pd.merge(left=survey_sub, right=species_sub, left_on='species_id', right
```

```
In [20]: merged_inner.shape
```

```
Out[20]: (10, 12)
```

```
In [21]: merged_inner
```

```
Out[21]:
```

	record_id	month	day	year	plot_id	species_id	sex	hindfoot_length	weight	genus
0	1	7	16	1977	2	NL	M	32.0	NaN	Neotoma
1	2	7	16	1977	3	NL	M	33.0	NaN	Neotoma
2	3	7	16	1977	2	DM	F	37.0	NaN	Dipodomys
3	4	7	16	1977	7	DM	M	36.0	NaN	Dipodomys
4	5	7	16	1977	3	DM	M	35.0	NaN	Dipodomys
5	8	7	16	1977	1	DM	M	37.0	NaN	Dipodomys
6	9	7	16	1977	1	DM	F	34.0	NaN	Dipodomys
7	6	7	16	1977	1	PF	M	14.0	NaN	Perognathus
8	10	7	16	1977	6	PF	F	20.0	NaN	Perognathus
9	7	7	16	1977	2	PE	F	NaN	NaN	Peromyscus

```
In [22]: merged_left = pd.merge(left=survey_sub, right=species_sub, how='left', left_on='species_id',
merged_left
```

```
Out[22]:
```

	record_id	month	day	year	plot_id	species_id	sex	hindfoot_length	weight	genus
0	1	7	16	1977	2	NL	M	32.0	NaN	Neotoma
1	2	7	16	1977	3	NL	M	33.0	NaN	Neotoma
2	3	7	16	1977	2	DM	F	37.0	NaN	Dipodomys
3	4	7	16	1977	7	DM	M	36.0	NaN	Dipodomys
4	5	7	16	1977	3	DM	M	35.0	NaN	Dipodomys
5	6	7	16	1977	1	PF	M	14.0	NaN	Perognathus
6	7	7	16	1977	2	PE	F	NaN	NaN	Peromyscus
7	8	7	16	1977	1	DM	M	37.0	NaN	Dipodomys
8	9	7	16	1977	1	DM	F	34.0	NaN	Dipodomys
9	10	7	16	1977	6	PF	F	20.0	NaN	Perognathus


```
In [24]: merged_left[merged_left['genus'].isna()]
```

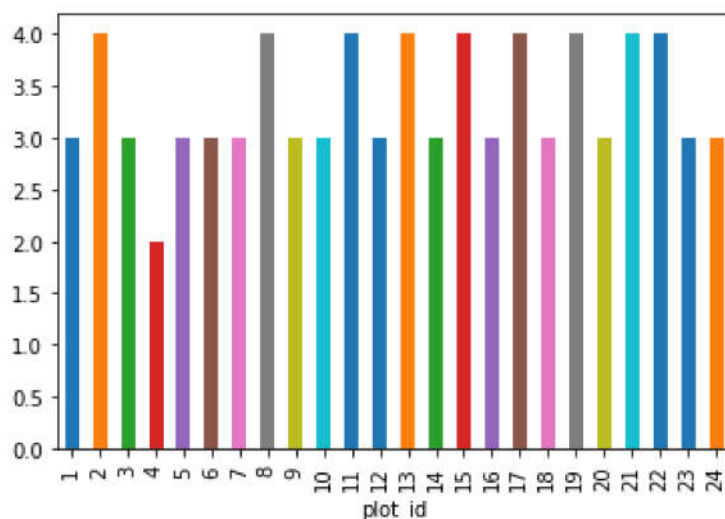
```
Out[24]:
```

```
record_id  month  day  year  plot_id  species_id  sex  hindfoot_length  weight  genus  species
```

```
In [25]: merged_left = pd.merge(left=surveys_df, right=species_df, how='left', on="species_id")
```

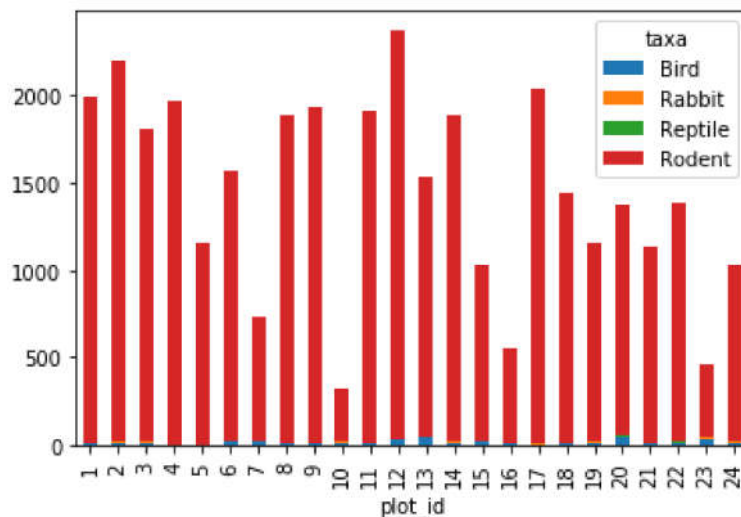
```
In [28]: merged_left.groupby(["plot_id"])["taxa"].nunique().plot(kind='bar')
```

```
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x2912ea09780>
```



```
In [29]: merged_left.groupby(["plot_id", "taxa"]).count()["record_id"].unstack().plot(kind='bar')
```

```
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x2912fb68860>
```



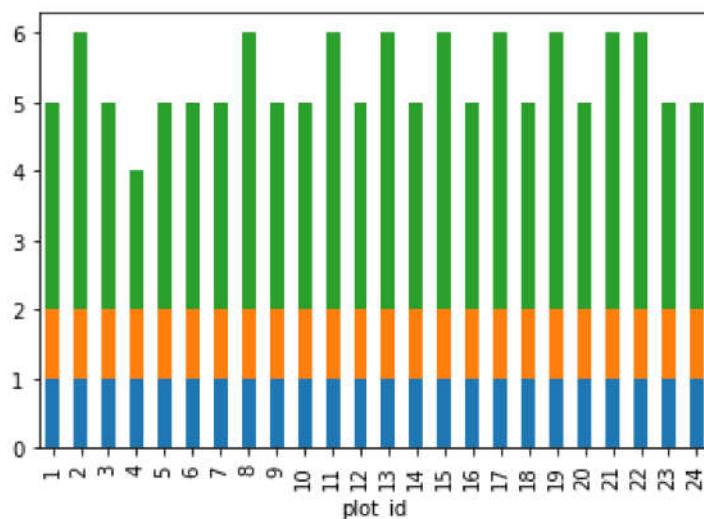
```
In [33]: merged_left.loc[merged_left["sex"].isnull(), "sex"] = 'M|F'
```

```
In [34]: ntaxa_sex_site = merged_left.groupby(["plot_id", "sex"])["taxa"].nunique().reset_index(1)
```

```
In [35]: ntaxa_sex_site = ntaxa_sex_site.pivot_table(values="taxa", columns="sex", index=ntaxa_s
```

```
In [36]: ntaxa_sex_site.plot(kind="bar", legend=False, stacked=True)
```

```
Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x2912e98c390>
```



```
In [42]: plot_info = pd.read_csv("plots.csv")
plot_info.groupby("plot_type").count()
```

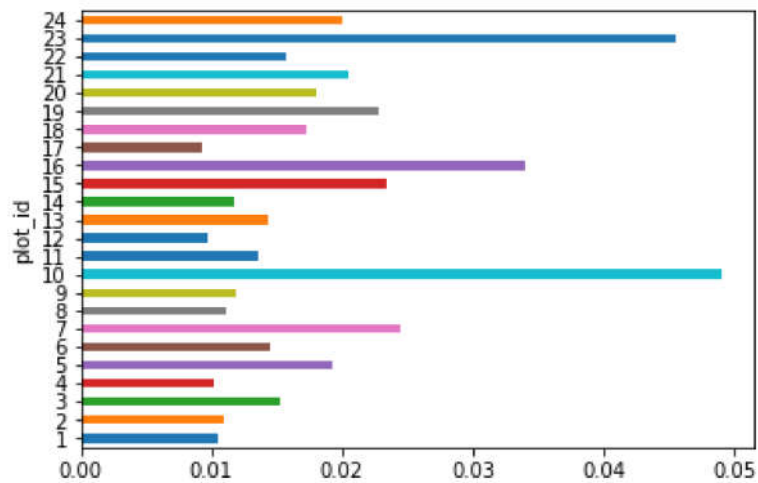
```
Out[42]:
```

	plot_id
plot_type	
Control	8
Long-term Krat Exclosure	4
Rodent Exclosure	6
Short-term Krat Exclosure	4
Spectab exclosure	2

```
In [44]: merged_site_type = pd.merge(merged_left, plot_info, on='plot_id')
# For each plot, get the number of species for each plot
nspecies_site = merged_site_type.groupby(["plot_id"])["species"].nunique().rename("nspe")
# For each plot, get the number of individuals
nindividuals_site = merged_site_type.groupby(["plot_id"]).count()["record_id"].rename("nindiv")
# combine the two series
diversity_index = pd.concat([nspecies_site, nindividuals_site], axis=1)
# calculate the diversity index
diversity_index['diversity'] = diversity_index['nspecies']/diversity_index['nindiv']
```

```
In [46]: diversity_index['diversity'].plot(kind="barh")
```

```
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x2912feb7fd0>
```



```
In [ ]:
```