Introduction

Maps allow us to transform data in a DataFrame or Series one value at a time for an entire column. However, often we want to group our data, and then do something specific to the group the data is in.

As you'll learn, we do this with the `groupby()` operation. We'll also cover some additional topics, such as more complex ways to index your DataFrames, along with how to sort your data.

Understanding Groupby¶

ฟังก์ชัน groupby() ใน Pandas จะแบ่งข้อมูลทั้งหมดจากชุดข้อมูลออกเป็นหมวดหมู่หรือกลุ่มต่างๆ ทำให้สามารถ วิเคราะห์ข้อมูล ตามกลุ่มต่างๆ ได้อย่างยืดหยุ่น

Here's a super simple dataframe to illustrate some examples. We'll be grouping the data by the "animal" column where there are four categories of animals:

- alligators
- cats
- snakes
- hamsters

In [1]:
```
import numpy as np
import pandas as pd
import random

# Random pets column
pet_list = ["cat", "hamster", "alligator", "snake"]
pet = [random.choice(pet_list) for i in range(1,15)]

# Random weight of animal column
weight = [random.choice(range(5,15)) for i in range(1,15)]

# Random length of animals column
length = [random.choice(range(1,10)) for i in range(1,15)]

# random age of the animals column
age = [random.choice(range(1,15)) for i in range(1,15)]

# Put everyhting into a dataframe
df = pd.DataFrame()
df["animal"] = pet
df["age"] = age
df["weight"] = weight
df["length"] = length
```

```
# make a groupby object
animal_groups = df.groupby("animal")
```

In [2]:
```
df
```

Out[2]:

	animal	age	weight	length
0	alligator	1	11	2
1	alligator	12	13	6
2	cat	6	14	7
3	alligator	2	9	7
4	alligator	13	9	1
5	cat	14	14	5
6	hamster	14	5	1
7	hamster	3	13	9
8	snake	3	14	7
9	alligator	9	9	1
10	hamster	13	14	1
11	snake	6	14	8
12	cat	13	13	9
13	alligator	7	7	3

- เราสามารถถามเกี่ยวกับข้อมูลสัตว์ได้ก็คือ
- หากต้องการหาค่าเฉลี่ย(**mean**)ของน้ำหนักของสัตว์แต่ละประเภท เราจะจัดกลุ่มสัตว์ตามประเภทของสัตว์ จากนั้นจึงใช้ฟังก์ชันค่าเฉลี่ย
- เราสามารถใช้ฟังก์ชันอื่นๆ ได้เช่นกัน
- เราสามารถใช้ "**sum**" เพื่อหาผลรวมน้ำหนักทั้งหมด
- "**min**" เพื่อค้นหาน้ำหนักต่ำสุด
- "**max**" เพื่อค้นหาน้ำหนักสูงสุด
- หรือ "**count**" เพื่อค้นหาจำนวนสัตว์แต่ละประเภท

Summary statistics	Numpy operations	More complex operations
mean	np.mean	.agg()
median	np.min	agg(["mean", "median"])
min	np.max	agg(custom_function())
max	np.sum	
sum	np.product	
describe		
count or size		

These two lines of code group the animals then apply the mean function to the weight column.

In [3]:
```
# Group by animal category
animal_groups = df.groupby("animal")
# Apply mean function to wieght column
animal_groups['weight'].mean()
```

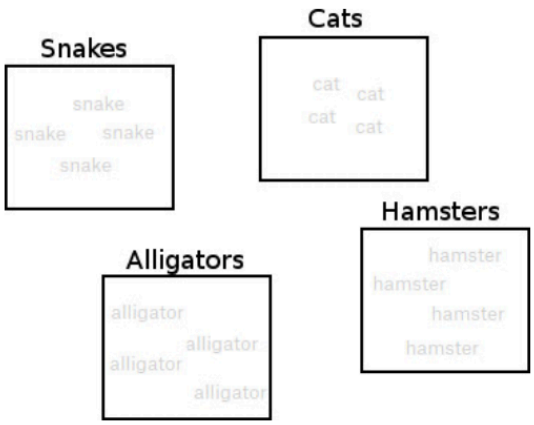
Out[3]:
```
animal
alligator     9.666667
cat          13.666667
hamster      10.666667
snake        14.000000
Name: weight, dtype: float64
```

Here's what happens when you run that code:

1. Group the unique values from the animal column

	animal	age	weight	length
0	snake	4	13	4
1	alligator	3	7	8
2	alligator	8	11	7
3	alligator	10	9	6
4	hamster	7	8	7
5	snake	11	6	7
6	hamster	14	13	2
7	hamster	13	13	5
8	hamster	2	11	9
9	snake	12	13	2
10	cat	8	11	5
11	cat	4	12	7
12	cat	3	11	9
13	alligator	7	9	2

2. Now there's a bucket for each group

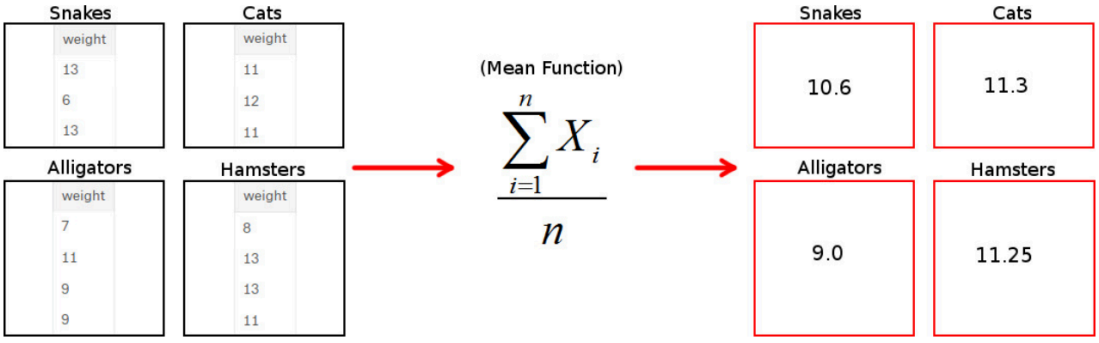


Loading [MathJax]/extensions/Safe.js

3. Toss the other data into the buckets

	animal	age	weight	length
0	snake	4	13	4
1	alligator	3	7	8
2	alligator	8	11	7
3	alligator	10	9	6
4	hamster	7	8	7
5	snake	11	6	7
6	hamster	14	13	2
7	hamster	13	13	5
8	hamster	2	11	9
9	snake	12	13	2
10	cat	8	11	5
11	cat	4	12	7
12	cat	3	11	9
13	alligator	7	9	2

4. Apply a function on the weight column of each bucket



In [4]:
```
df.groupby("animal").size()
```

Out[4]:
```
animal
alligator    6
cat          3
hamster      3
snake        2
dtype: int64
```

In [5]:
```
df.groupby("animal").count()
```

Out[5]:

animal	age	weight	length
alligator	6	6	6
cat	3	3	3
hamster	3	3	3
snake	2	2	2

In [6]:
```
df.groupby("animal")["weight"].count()
```

Out[6]:
```
animal
alligator    6
cat          3
hamster      3
snake        2
Name: weight, dtype: int64
```

In [7]:
```
df.groupby("weight")["animal"].count()
```

Out[7]:
```
weight
5     1
7     1
9     3
11    1
13    3
14    5
Name: animal, dtype: int64
```

In [8]:
```
df.groupby("weight")["animal"].max()
```

Out[8]:
```
weight
5       hamster
7     alligator
9     alligator
11    alligator
13      hamster
14        snake
Name: animal, dtype: object
```

Next Example

In [9]:
```
import pandas as pd

reviews = pd.read_csv("datasets/winemag-data-52.csv", index_col=0)
reviews
```

Out[9]:

	country	description	designation	points	price	province	region_1	reg
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	90	NaN	Sicily & Sardinia	Etna	
1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	
2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willa
3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	90	13.0	Michigan	Lake Michigan Shore	
4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	95	65.0	Oregon	Willamette Valley	Willa
5	Spain	Blackberry and raspberry aromas show a typical...	Ars In Vitro	87	15.0	Northern Spain	Navarra	
6	Italy	Here's a bright, informal red that opens with ...	Belsito	92	16.0	Sicily & Sardinia	Vittoria	
7	France	This dry and restrained wine offers spice in p...	NaN	90	24.0	Alsace	Alsace	
8	Germany	Savory dried thyme notes accent sunnier flavor...	Shine	87	12.0	Rheinhessen	NaN	
9	France	This has great depth of flavor with its fresh ...	Les Natures	92	27.0	Alsace	Alsace	
10	US	Soft, supple plum envelopes an	Mountain Cuvée	87	19.0	California	Napa Valley	

	country	description	designation	points	price	province	region_1	reg
		oaky structure ...						
11	France	This is a dry wine, very spicy, with a tight, ...	NaN	87	30.0	Alsace	Alsace	
12	US	Slightly reduced, this wine offers a chalky, t...	NaN	94	34.0	California	Alexander Valley	S
13	Italy	This is dominated by oak and oak-driven aromas...	Rosso	95	NaN	Sicily & Sardinia	Etna	
14	US	Building on 150 years and six generations of w...	NaN	92	12.0	California	Central Coast	C
15	Germany	Zesty orange peels and apple notes abound in t...	Devon	87	24.0	Mosel	NaN	
16	Argentina	Baked plum, molasses, balsamic vinegar and che...	Felix	87	30.0	Other	Cafayate	
17	Argentina	Raw black-cherry aromas are direct and simple ...	Winemaker Selection	87	13.0	Mendoza Province	Mendoza	
18	Spain	Desiccated blackberry, leather, charred wood a...	Vendimia Seleccionada Finca Valdelayegua Singl...	87	28.0	Northern Spain	Ribera del Duero	
19	US	Red fruit aromas pervade on the nose, with cig...	NaN	90	32.0	Virginia	Virginia	
20	US	Ripe aromas of dark berries mingle with ample ...	Vin de Maison	87	23.0	Virginia	Virginia	

	country	description	designation	points	price	province	region_1	reg
21	US	A sleek mix of tart berry, stem and herb, alon...	NaN	92	20.0	Oregon	Oregon	C
22	Italy	Delicate aromas recall white flower and citrus...	Ficiligno	92	19.0	Sicily & Sardinia	Sicilia	
23	US	This wine from the Geneseo district offers aro...	Signature Selection	94	22.0	California	Paso Robles	C
24	Italy	Aromas of prune, blackcurrant, toast and oak c...	Aynat	90	35.0	Sicily & Sardinia	Sicilia	
25	US	Oak and earth intermingle around robust aromas...	King Ridge Vineyard	95	69.0	California	Sonoma Coast	S
26	Italy	Pretty aromas of yellow flower and stone fruit...	Dalila	95	13.0	Sicily & Sardinia	Terre Siciliane	
27	Italy	Aromas recall ripe dark berry, toast and a whi...	NaN	87	10.0	Sicily & Sardinia	Terre Siciliane	
28	Italy	Aromas suggest mature berry, scorched earth, a...	Mascaria Barricato	90	17.0	Sicily & Sardinia	Cerasuolo di Vittoria	
29	US	Clarksburg is becoming a haven for Chenin Blan...	NaN	86	16.0	California	Clarksburg	C
30	France	Red cherry fruit comes laced with light tannin...	Nouveau	86	NaN	Beaujolais	Beaujolais-Villages	

	country	description	designation	points	price	province	region_1	reg
31	Italy	Merlot and Nero d'Avola form the base for this...	Calanìca Nero d'Avola-Merlot	86	NaN	Sicily & Sardinia	Sicilia	
32	Italy	Part of the extended Calanìca series, this Gri...	Calanìca Grillo-Viognier	90	NaN	Sicily & Sardinia	Sicilia	
33	US	Rustic and dry, this has flavors of berries, c...	Puma Springs Vineyard	86	50.0	California	Dry Creek Valley	S
34	US	This shows a tart, green gooseberry flavor tha...	NaN	94	20.0	California	Sonoma Valley	S
35	US	As with many of the Erath 2010 vineyard design...	Hyland	95	50.0	Oregon	McMinnville	Willa
36	Chile	White flower, lychee and apple aromas carry th...	Estate	86	15.0	Colchagua Valley	NaN	
37	Italy	This concentrated Cabernet offers aromas of cu...	Missoni	95	21.0	Sicily & Sardinia	Sicilia	
38	Italy	Inky in color, this wine has plump aromas of r...	I Tratturi	92	11.0	Southern Italy	Puglia	
39	Italy	Part of the natural wine movement, this wine i...	Purato Made With Organic Grapes	86	12.0	Sicily & Sardinia	Sicilia	
40	Italy	Catarratto is one of Sicily's most widely farm...	NaN	86	17.0	Sicily & Sardinia	Sicilia	
41	US	A stiff, tannic wine, this	NaN	92	22.0	Oregon	Willamette Valley	Willa

	country	description	designation	points	price	province	region_1	reg
		slowly opens and br...						
42	France	This is a festive wine, with soft, ripe fruit ...	Nouveau	86	9.0	Beaujolais	Beaujolais	
43	US	The clean, brisk mouthfeel gives this slightly...	NaN	86	14.0	California	Paso Robles	C
44	Chile	A berry aroma comes with cola and herb notes. ...	NaN	86	9.0	Maule Valley	NaN	
45	US	Right out of the starting blocks this is an oa...	#SocialSecret	90	40.0	Virginia	Virginia	
46	Italy	Spicy, fresh and clean, this would pair with f...	Sallier de la Tour	92	13.0	Sicily & Sardinia	Sicilia	
47	US	This is a sweet wine with flavors of white sug...	NaN	90	13.0	California	Lake County	
48	US	This bottling resembles the New Zealand paradi...	NaN	95	16.0	Virginia	Monticello	
49	France	Soft and fruity, this is a generous, ripe wine...	Eté Indien	86	14.0	Beaujolais	Brouilly	
50	Italy	This blend of Nero d'Avola and Syrah opens wit...	Scialo	86	NaN	Sicily & Sardinia	Sicilia	
51	Chile	This is much different than Casa Silva's 2009 ...	Gran Reserva	85	22.0	Colchagua Valley	NaN	

In [10]:
```
#เช็คว่า DataFrame มีคอลัมน์อะไรบ้าง
reviews.columns
```

Out[10]:
```
Index(['country', 'description', 'designation', 'points', 'price', 'provinc
e',
       'region_1', 'region_2', 'taster_name', 'taster_twitter_handle', 'tit
le',
       'variety', 'winery'],
      dtype='object')
```

In [11]:
```
#แสดงผลทางสถิติโดยยึด country เป็นหลัก
reviews.groupby('country').describe()
```

Out[11]:

	points									
	count	mean	std	min	25%	50%	75%	max	count	m
country										
Argentina	2.0	87.000000	0.000000	87.0	87.00	87.0	87.00	87.0	2.0	21.500
Chile	3.0	85.666667	0.577350	85.0	85.50	86.0	86.00	86.0	3.0	15.333
France	6.0	87.833333	2.562551	86.0	86.00	86.5	89.25	92.0	5.0	20.800
Germany	2.0	87.000000	0.000000	87.0	87.00	87.0	87.00	87.0	2.0	18.000
Italy	16.0	90.250000	3.296463	86.0	86.75	90.0	92.00	95.0	11.0	16.727
Portugal	1.0	87.000000	NaN	87.0	87.00	87.0	87.00	87.0	1.0	15.000
Spain	2.0	87.000000	0.000000	87.0	87.00	87.0	87.00	87.0	2.0	21.500
US	20.0	90.850000	3.407036	86.0	87.00	91.0	94.00	95.0	20.0	28.200

In [12]:
```
reviews.groupby('country').size()
```

Out[12]:
```
country
Argentina     2
Chile         3
France        6
Germany       2
Italy        16
Portugal      1
Spain         2
US           20
dtype: int64
```

In [13]:
```
#นับจำนวนทุกคอลัมน์โดยยึด country เป็นหลัก
reviews.groupby('country').count()
```

Out[13]:

country	description	designation	points	price	province	region_1	region_2	taste
Argentina	2	2	2	2	2	2	0	
Chile	3	2	3	3	3	0	0	
France	6	4	6	5	6	6	0	
Germany	2	2	2	2	2	0	0	
Italy	16	14	16	11	16	16	0	
Portugal	1	1	1	1	1	0	0	
Spain	2	2	2	2	2	2	0	
US	20	9	20	20	20	20	14	

In [14]:
```
reviews.groupby('points').size()
```

Out[14]:
```
points
85     1
86    12
87    12
90     9
92     8
94     3
95     7
dtype: int64
```

In [15]:
```
#นับจำนวนทุกคอลัมน์โดยยึด points เป็นหลัก
reviews.groupby('points').count()
```

Out[15]:

points	country	description	designation	price	province	region_1	region_2	taster_
85	1	1	1	1	1	0	0	
86	12	12	8	9	12	10	3	
87	12	12	9	12	12	9	2	
90	9	9	6	7	9	9	0	
92	8	8	5	8	8	8	3	
94	3	3	1	3	3	3	3	
95	7	7	6	6	7	7	3	

In [16]:
```
#นับจำนวนคอลัมน์ price โดยยึด points เป็นหลัก
reviews.groupby('points').price.count()
```

Out[16]:
```
points
85     1
86     9
87    12
90     7
92     8
94     3
95     6
Name: price, dtype: int64
```

In [17]:
```
#นับจำนวนคอลัมน์ country โดยยึด points เป็นหลัก
reviews.groupby('points').country.count()
```

Out[17]:
```
points
85     1
86    12
87    12
90     9
92     8
94     3
95     7
Name: country, dtype: int64
```

In [18]:
```
#นับจำนวนคอลัมน์ country แยกแต่ละ value(ประเทศ) โดยยึด points เป็นหลัก
reviews.groupby('points').country.value_counts()
```

Out[18]:
```
points  country
85      Chile        1
86      Italy        4
        France       3
        US           3
        Chile        2
87      US           3
        Spain        2
        Germany      2
        Argentina    2
        Portugal     1
        Italy        1
        France       1
90      US           4
        Italy        4
        France       1
92      Italy        4
        US           3
        France       1
94      US           3
95      US           4
        Italy        3
Name: count, dtype: int64
```

In [19]:
```
#นับจำนวนคอลัมน์ price แยกแต่ละ value(ราคา) โดยยึด points เป็นหลัก
reviews.groupby('points').price.value_counts()
```

Out[19]:

points	price	
85	22.0	1
86	14.0	2
	9.0	2
	12.0	1
	15.0	1
	16.0	1
	17.0	1
	50.0	1
87	15.0	2
	30.0	2
	12.0	1
	10.0	1
	13.0	1
	14.0	1
	19.0	1
	23.0	1
	24.0	1
	28.0	1
90	13.0	2
	40.0	1
	24.0	1
	35.0	1
	32.0	1
	17.0	1
92	12.0	1
	13.0	1
	16.0	1
	19.0	1
	20.0	1
	22.0	1
	27.0	1
	11.0	1
94	20.0	1
	22.0	1
	34.0	1
95	13.0	1
	16.0	1
	21.0	1
	50.0	1
	65.0	1
	69.0	1

Name: count, dtype: int64

`groupby()` created a group of reviews which allotted the same point values to the given wines. Then, for each of these groups, we grabbed the `points()` column and counted how many times it appeared. `value_counts()` is just a shortcut to this `groupby()` operation.

We can use any of the summary functions we've used before with this data. For example, to get the cheapest wine in each point value category, we can do the following:

In [20]:
```
#หาราคาถูกที่สุดในคอลัมน์ price โดยยึด points เป็นหลัก
reviews.groupby('points').price.min()
```

Out[20]:
```
points
85    22.0
86     9.0
87    10.0
90    13.0
92    11.0
94    20.0
95    13.0
Name: price, dtype: float64
```

You can think of each group we generate as being a slice of our DataFrame containing only data with values that match. This DataFrame is accessible to us directly using the `apply()` method, and we can then manipulate the data in any way we see fit. For example, here's one way of selecting the name of the first wine reviewed from each winery in the dataset:

การใช้ apply() กับ groupby()

เราสามารถใช้ apply() เพื่อทำงานกับข้อมูลในแต่ละกลุ่มได้อย่างยืดหยุ่น

In [21]:
```
#แสดงผลชื่อแรกในคอลัมน์ title โดยยึด points เป็นหลัก
reviews.groupby('points').apply(lambda df: df.title.iloc[0])
```

Out[21]:
```
points
85    Casa Silva 2008 Gran Reserva Petit Verdot (Col...
86    Clarksburg Wine Company 2010 Chenin Blanc (Cla...
87        Quinta dos Avidagos 2011 Avidagos Red (Douro)
90                    Nicosia 2013 Vulkà Bianco  (Etna)
92     Terre di Giurfo 2013 Belsito Frappato (Vittoria)
94    Louis M. Martini 2012 Cabernet Sauvignon (Alex...
95    Sweet Cheeks 2012 Vintner's Reserve Wild Child...
dtype: object
```

In [22]:
```
#แสดงผลชื่อแรกในคอลัมน์ title โดยยึด winery เป็นหลัก(เลือกชื่อไวน์รายการแรกที่รีวิวจากแต่ละโ
reviews.groupby('winery').apply(lambda df: df.title.iloc[0])
```

Out[22]:

winery	
Acrobat (Oregon)	Acrobat 2013 Pinot Noir
Baglio di Pianetto (Sicilia)	Baglio di Pianetto 2007 Ficiligno White
Bianchi (Paso ...	Bianchi 2011 Signature Selection Merlot
Canicattì (Sicilia)	Canicattì 2009 Aynat Nero d'Avola
Casa Silva t (Col...	Casa Silva 2008 Gran Reserva Petit Verdo
Castello di Amorosa yard P...	Castello di Amorosa 2011 King Ridge Vine
Clarksburg Wine Company c (Cla...	Clarksburg Wine Company 2010 Chenin Blan
Domaine de la Madone jolais...	Domaine de la Madone 2012 Nouveau (Beau
Duca di Salaparuta d'Avola-...	Duca di Salaparuta 2010 Calanìca Nero
Envolve (Dry Cr...	Envolve 2010 Puma Springs Vineyard Red
Erath innville)	Erath 2010 Hyland Pinot Noir (McM
Estampa (Colch...	Estampa 2011 Estate Viognier-Chardonnay
Felix Lavaque (Cafayate)	Felix Lavaque 2010 Felix Malbec
Feudi del Pisciotto t Sauv...	Feudi del Pisciotto 2010 Missoni Caberne
Feudi di San Marzano mitivo...	Feudi di San Marzano 2011 I Tratturi Pri
Feudo Montoni (Sicilia)	Feudo Montoni 2011 Catarratto
Feudo di Santa Tresa th Org...	Feudo di Santa Tresa 2011 Purato Made Wi
Gaucho Andino albec ...	Gaucho Andino 2011 Winemaker Selection M
Hawkins Cellars tte Va...	Hawkins Cellars 2009 Pinot Noir (Willame
Heinz Eifel heinhe...	Heinz Eifel 2013 Shine Gewürztraminer (R
Henry Fessy aujolais)	Henry Fessy 2012 Nouveau (Be
Jean-Baptiste Adam t Gris...	Jean-Baptiste Adam 2012 Les Natures Pino
Kirkland Signature aberne...	Kirkland Signature 2011 Mountain Cuvée C
Leon Beyer (Alsace)	Leon Beyer 2012 Gewurztraminer
Louis M. Martini (Alex...	Louis M. Martini 2012 Cabernet Sauvignon
Masseria Setteporte o (Etna)	Masseria Setteporte 2012 Ross
Mirassou	Mirassou 2012 Chardonnay (Centr
Nicosia	Nicosia 2013 Vulkà Bianc

```
o  (Etna)
Pradorey                                Pradorey 2010 Vendimia Seleccionada Finc
a Vald...
Quinta dos Avidagos                         Quinta dos Avidagos 2011 Avidagos Re
d (Douro)
Quiévremont                                          Quiévremont 2012 Meritage
(Virginia)
Rainstorm                                   Rainstorm 2013 Pinot Gris (Willamett
e Valley)
Richard Böcking                               Richard Böcking 2013 Devon Rieslin
g (Mosel)
Robert Hall                                Robert Hall 2011 Sauvignon Blanc (Pas
o Robles)
St. Julian                              St. Julian 2013 Reserve Late Harvest Rie
sling ...
Stemmari                                     Stemmari 2013 Dalila White (Terre S
iciliane)
Sundance                                             Sundance 2011 Merlot (Maul
e Valley)
Sweet Cheeks                            Sweet Cheeks 2012 Vintner's Reserve Wild
Child...
Tandem                                  Tandem 2011 Ars In Vitro Tempranillo-Mer
lot (N...
Tarara                                          Tarara 2010 #SocialSecret Red
(Virginia)
Tasca d'Almerita                        Tasca d'Almerita 2011 Sallier de la Tour
Inzol...
Terre di Giurfo                          Terre di Giurfo 2013 Belsito Frappato
(Vittoria)
The White Knight                             The White Knight 2011 Riesling (Lak
e County)
Trimbach                                            Trimbach 2012 Gewurztraminer
(Alsace)
Trump                                           Trump 2011 Sauvignon Blanc (Mo
nticello)
Vignerons de Bel Air                     Vignerons de Bel Air 2011 Eté Indien
(Brouilly)
Viticultori Associati Canicatti    Viticultori Associati Canicatti 2008 Sci
alo Re...
dtype: object
```

In [23]:
```
reviews.groupby('winery').title.count()
```

Out[23]:
```
winery
Acrobat                          1
Baglio di Pianetto               1
Bianchi                          1
Canicattì                        1
Casa Silva                       1
Castello di Amorosa              1
Clarksburg Wine Company          1
Domaine de la Madone             1
Duca di Salaparuta               2
Envolve                          2
Erath                            1
Estampa                          1
Felix Lavaque                    1
Feudi del Pisciotto              1
Feudi di San Marzano             1
Feudo Montoni                    1
Feudo di Santa Tresa             1
Gaucho Andino                    1
Hawkins Cellars                  1
Heinz Eifel                      1
Henry Fessy                      1
Jean-Baptiste Adam               1
Kirkland Signature               1
Leon Beyer                       1
Louis M. Martini                 1
Masseria Setteporte              1
Mirassou                         1
Nicosia                          1
Pradorey                         1
Quinta dos Avidagos              1
Quiévremont                      2
Rainstorm                        1
Richard Böcking                  1
Robert Hall                      1
St. Julian                       1
Stemmari                         2
Sundance                         1
Sweet Cheeks                     1
Tandem                           1
Tarara                           1
Tasca d'Almerita                 1
Terre di Giurfo                  2
The White Knight                 1
Trimbach                         1
Trump                            1
Vignerons de Bel Air             1
Viticultori Associati Canicatti  1
Name: title, dtype: int64
```

For even more fine-grained control, you can also group by more than one column. For an example, here's how we would pick out the best wine by country *and* province:

In [24]:
```
# เลือกไวน์ที่คะแนนดีที่สุดตามประเทศ
reviews.groupby(['country']).apply(lambda df: df.loc[df.points.idxmax()])
```

Out[24]:

country	country	description	designation	points	price	province	region_1
Argentina	Argentina	Baked plum, molasses, balsamic vinegar and che...	Felix	87	30.0	Other	Cafayate
Chile	Chile	White flower, lychee and apple aromas carry th...	Estate	86	15.0	Colchagua Valley	NaN
France	France	This has great depth of flavor with its fresh ...	Les Natures	92	27.0	Alsace	Alsace
Germany	Germany	Savory dried thyme notes accent sunnier flavor...	Shine	87	12.0	Rheinhessen	NaN
Italy	Italy	This is dominated by oak and oak-driven aromas...	Rosso	95	NaN	Sicily & Sardinia	Etna
Portugal	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN
Spain	Spain	Blackberry and raspberry aromas show a typical...	Ars In Vitro	87	15.0	Northern Spain	Navarra
US	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	95	65.0	Oregon	Willamette Valley

In [25]:
```
#เลือกไวน์ที่คะแนนดีที่สุดตามจังหวัด
reviews.groupby(['province']).apply(lambda df: df.loc[df.points.idxmax()])
```

Out[25]:

province	country	description	designation	points	price	province	region_
Alsace	France	This has great depth of flavor with its fresh ...	Les Natures	92	27.0	Alsace	Alsac
Beaujolais	France	Red cherry fruit comes laced with light tannin...	Nouveau	86	NaN	Beaujolais	Beaujolais Village
California	US	Oak and earth intermingle around robust aromas...	King Ridge Vineyard	95	69.0	California	Sonom Coas
Colchagua Valley	Chile	White flower, lychee and apple aromas carry th...	Estate	86	15.0	Colchagua Valley	Nal
Douro	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	Nal
Maule Valley	Chile	A berry aroma comes with cola and herb notes. ...	NaN	86	9.0	Maule Valley	Nal
Mendoza Province	Argentina	Raw black-cherry aromas are direct and simple ...	Winemaker Selection	87	13.0	Mendoza Province	Mendoz
Michigan	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	90	13.0	Michigan	Lak Michiga Shor
Mosel	Germany	Zesty orange peels and apple notes abound in t...	Devon	87	24.0	Mosel	Nal

province	country	description	designation	points	price	province	region_
Northern Spain	Spain	Blackberry and raspberry aromas show a typical...	Ars In Vitro	87	15.0	Northern Spain	Navarr
Oregon	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	95	65.0	Oregon	Willamett Valle
Other	Argentina	Baked plum, molasses, balsamic vinegar and che...	Felix	87	30.0	Other	Cafayat
Rheinhessen	Germany	Savory dried thyme notes accent sunnier flavor...	Shine	87	12.0	Rheinhessen	Nal
Sicily & Sardinia	Italy	This is dominated by oak and oak-driven aromas...	Rosso	95	NaN	Sicily & Sardinia	Etn
Southern Italy	Italy	Inky in color, this wine has plump aromas of r...	I Tratturi	92	11.0	Southern Italy	Pugli
Virginia	US	This bottling resembles the New Zealand paradi...	NaN	95	16.0	Virginia	Monticell

In [26]:
```
#สามารถจัดกลุ่มได้มากกว่าหนึ่งคอลัมน์
#ตัวอย่างเช่น เราจะเลือกไวน์ที่คะแนนดีที่สุดตามประเทศและจังหวัดได้อย่างไร:
reviews.groupby(['country', 'province']).apply(lambda df: df.loc[df.points.i
```

Loading [MathJax]/extensions/Safe.js

Out[26]:

country	province	country	description	designation	points	price	province
Argentina	**Mendoza Province**	Argentina	Raw black-cherry aromas are direct and simple ...	Winemaker Selection	87	13.0	Mendoza Province
	Other	Argentina	Baked plum, molasses, balsamic vinegar and che...	Felix	87	30.0	Other
Chile	**Colchagua Valley**	Chile	White flower, lychee and apple aromas carry th...	Estate	86	15.0	Colchagua Valley
	Maule Valley	Chile	A berry aroma comes with cola and herb notes. ...	NaN	86	9.0	Maule Valley
France	**Alsace**	France	This has great depth of flavor with its fresh ...	Les Natures	92	27.0	Alsace
	Beaujolais	France	Red cherry fruit comes laced with light tannin...	Nouveau	86	NaN	Beaujolais
Germany	**Mosel**	Germany	Zesty orange peels and apple notes abound in t...	Devon	87	24.0	Mosel
	Rheinhessen	Germany	Savory dried thyme notes accent sunnier flavor...	Shine	87	12.0	Rheinhessen
Italy	**Sicily & Sardinia**	Italy	This is dominated by oak and	Rosso	95	NaN	Sicily & Sardinia

country	province	country	description	designation	points	price	province
			oak-driven aromas...				
	Southern Italy	Italy	Inky in color, this wine has plump aromas of r...	I Tratturi	92	11.0	Southern Italy
Portugal	**Douro**	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro
Spain	**Northern Spain**	Spain	Blackberry and raspberry aromas show a typical...	Ars In Vitro	87	15.0	Northern Spain
US	**California**	US	Oak and earth intermingle around robust aromas...	King Ridge Vineyard	95	69.0	California
	Michigan	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	90	13.0	Michigan
	Oregon	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	95	65.0	Oregon
	Virginia	US	This bottling resembles the New Zealand paradi...	NaN	95	16.0	Virginia

Another `groupby()` method worth mentioning is `agg()`, which lets you run a bunch of different functions on your DataFrame simultaneously. For example, we can generate a simple statistical summary of the dataset as follows:

การใช้ agg() กับ groupby()

ฟังก์ชัน agg() (Aggregate functionคือฟังก์ชันสรุปผล เช่น len,min,max,sum,count,mean,median) ช่วยให้เราสามารถทำการวิเคราะห์หลายรูปแบบพร้อมกัน

In [27]:
```
#เราสามารถใช้คือ agg()
reviews.groupby(['country']).price.agg([len, min, max])
```

```
/var/folders/83/3fg00w111r7bf7rcsz4nznlh0000gn/T/ipykernel_21811/4213695002.
py:2: FutureWarning: The provided callable <built-in function min> is curren
tly using SeriesGroupBy.min. In a future version of pandas, the provided cal
lable will be used directly. To keep current behavior pass the string "min"
instead.
  reviews.groupby(['country']).price.agg([len, min, max])
/var/folders/83/3fg00w111r7bf7rcsz4nznlh0000gn/T/ipykernel_21811/4213695002.
py:2: FutureWarning: The provided callable <built-in function max> is curren
tly using SeriesGroupBy.max. In a future version of pandas, the provided cal
lable will be used directly. To keep current behavior pass the string "max"
instead.
  reviews.groupby(['country']).price.agg([len, min, max])
```

Out[27]:

country	len	min	max
Argentina	2	13.0	30.0
Chile	3	9.0	22.0
France	6	9.0	30.0
Germany	2	12.0	24.0
Italy	16	10.0	35.0
Portugal	1	15.0	15.0
Spain	2	15.0	28.0
US	20	12.0	69.0

Effective use of `groupby()` will allow you to do lots of really powerful things with your dataset.

Multi-indexes

In all of the examples we've seen thus far we've been working with DataFrame or Series objects with a single-label index. `groupby()` is slightly different in the fact that, depending on the operation we run, it will sometimes result in what is called a multi-index.

A multi-index differs from a regular index in that it has multiple levels. For example:

เมื่อค่าของคอลัมน์เดียวไม่เพียงพอที่จะระบุแถวได้อย่างชัดเจน (เช่น ระเบียนหลายรายการในวันที่เดียวกัน ที่หมายความว่าวันที่เพียงอย่างเดียวไม่เหมาะเป็นดัชนี) เมื่อข้อมูลมีลำดับชั้นเชิงตรรกะ ซึ่ง

หมายความว่ามีมิติหรือ "ระดับ" หลายระดับ นอกจากโครงสร้างแล้ว ดัชนีหลายดัชนียังช่วยให้เรียกค้นข้อมูลที่ซับซ้อนในหน่วยความจำได้ค่อนข้างง่าย

In [28]:
```
countries_reviewed = reviews.groupby(['country', 'province']).title.agg([ler
countries_reviewed
```

Out[28]:

		len
country	**province**	
Argentina	**Mendoza Province**	1
	Other	1
Chile	**Colchagua Valley**	2
	Maule Valley	1
France	**Alsace**	3
	Beaujolais	3
Germany	**Mosel**	1
	Rheinhessen	1
Italy	**Sicily & Sardinia**	15
	Southern Italy	1
Portugal	**Douro**	1
Spain	**Northern Spain**	2
US	**California**	10
	Michigan	1
	Oregon	5
	Virginia	4

In [29]:
```
countries_reviewed.index
```

Out[29]:
```
MultiIndex([('Argentina',  'Mendoza Province'),
            ('Argentina',             'Other'),
            (    'Chile',  'Colchagua Valley'),
            (    'Chile',      'Maule Valley'),
            (   'France',           'Alsace'),
            (   'France',        'Beaujolais'),
            (  'Germany',             'Mosel'),
            (  'Germany',       'Rheinhessen'),
            (    'Italy', 'Sicily & Sardinia'),
            (    'Italy',    'Southern Italy'),
            ( 'Portugal',             'Douro'),
            (    'Spain',    'Northern Spain'),
            (       'US',        'California'),
            (       'US',          'Michigan'),
            (       'US',            'Oregon'),
            (       'US',          'Virginia')],
           names=['country', 'province'])
```

In [30]:
```
type(countries_reviewed.index)
```

Out[30]:
```
pandas.core.indexes.multi.MultiIndex
```

Multi-indices have several methods for dealing with their tiered structure which are absent for single-level indices. They also require two levels of labels to retrieve a value. Dealing with multi-index output is a common "gotcha" for users new to pandas.

The use cases for a multi-index are detailed alongside instructions on using them in the MultiIndex / Advanced Selection section of the pandas documentation.

However, in general the multi-index method you will use most often is the one for converting back to a regular index, the `reset_index()` method:

In [31]:
```
countries_reviewed.reset_index()
```

Out[31]:

	country	province	len
0	Argentina	Mendoza Province	1
1	Argentina	Other	1
2	Chile	Colchagua Valley	2
3	Chile	Maule Valley	1
4	France	Alsace	3
5	France	Beaujolais	3
6	Germany	Mosel	1
7	Germany	Rheinhessen	1
8	Italy	Sicily & Sardinia	15
9	Italy	Southern Italy	1
10	Portugal	Douro	1
11	Spain	Northern Spain	2
12	US	California	10
13	US	Michigan	1
14	US	Oregon	5
15	US	Virginia	4

In [32]:
```
type(countries_reviewed.reset_index())
```

Out[32]:
```
pandas.core.frame.DataFrame
```

Sorting

Looking again at `countries_reviewed` we can see that grouping returns data in index order, not in value order. That is to say, when outputting the result of a `groupby`, the order of the rows is dependent on the values in the index, not in the data.

To get data in the order want it in we can sort it ourselves. The `sort_values()` method is handy for this.

In [33]:
```
countries_reviewed = countries_reviewed.reset_index()
countries_reviewed.sort_values(by='len')
```

Out[33]:

	country	province	len
0	Argentina	Mendoza Province	1
1	Argentina	Other	1
3	Chile	Maule Valley	1
6	Germany	Mosel	1
7	Germany	Rheinhessen	1
9	Italy	Southern Italy	1
10	Portugal	Douro	1
13	US	Michigan	1
2	Chile	Colchagua Valley	2
11	Spain	Northern Spain	2
4	France	Alsace	3
5	France	Beaujolais	3
15	US	Virginia	4
14	US	Oregon	5
12	US	California	10
8	Italy	Sicily & Sardinia	15

In [34]:
```
countries_reviewed.sort_values('len')
```

Out[34]:

	country	province	len
0	Argentina	Mendoza Province	1
1	Argentina	Other	1
3	Chile	Maule Valley	1
6	Germany	Mosel	1
7	Germany	Rheinhessen	1
9	Italy	Southern Italy	1
10	Portugal	Douro	1
13	US	Michigan	1
2	Chile	Colchagua Valley	2
11	Spain	Northern Spain	2
4	France	Alsace	3
5	France	Beaujolais	3
15	US	Virginia	4
14	US	Oregon	5
12	US	California	10
8	Italy	Sicily & Sardinia	15

`sort_values()` defaults to an ascending sort, where the lowest values go first. However, most of the time we want a descending sort, where the higher numbers go first. That goes thusly:

In [35]:
```
countries_reviewed.sort_values(by='len').iloc[::-1]
```

Out[35]:

	country	province	len
8	Italy	Sicily & Sardinia	15
12	US	California	10
14	US	Oregon	5
15	US	Virginia	4
5	France	Beaujolais	3
4	France	Alsace	3
11	Spain	Northern Spain	2
2	Chile	Colchagua Valley	2
13	US	Michigan	1
10	Portugal	Douro	1
9	Italy	Southern Italy	1
7	Germany	Rheinhessen	1
6	Germany	Mosel	1
3	Chile	Maule Valley	1
1	Argentina	Other	1
0	Argentina	Mendoza Province	1

In [36]:
```
countries_reviewed.sort_values(by='len', ascending=False)
```

Out[36]:

	country	province	len
8	Italy	Sicily & Sardinia	15
12	US	California	10
14	US	Oregon	5
15	US	Virginia	4
4	France	Alsace	3
5	France	Beaujolais	3
2	Chile	Colchagua Valley	2
11	Spain	Northern Spain	2
0	Argentina	Mendoza Province	1
1	Argentina	Other	1
3	Chile	Maule Valley	1
6	Germany	Mosel	1
7	Germany	Rheinhessen	1
9	Italy	Southern Italy	1
10	Portugal	Douro	1
13	US	Michigan	1

To sort by index values, use the companion method `sort_index()`. This method has the same arguments and default order:

In [37]:
```
countries_reviewed.sort_index()
```

Out[37]:

	country	province	len
0	Argentina	Mendoza Province	1
1	Argentina	Other	1
2	Chile	Colchagua Valley	2
3	Chile	Maule Valley	1
4	France	Alsace	3
5	France	Beaujolais	3
6	Germany	Mosel	1
7	Germany	Rheinhessen	1
8	Italy	Sicily & Sardinia	15
9	Italy	Southern Italy	1
10	Portugal	Douro	1
11	Spain	Northern Spain	2
12	US	California	10
13	US	Michigan	1
14	US	Oregon	5
15	US	Virginia	4

Finally, know that you can sort by more than one column at a time:

In [38]:
```
countries_reviewed.sort_values(by=['country', 'len'])
```

Out[38]:

	country	province	len
0	Argentina	Mendoza Province	1
1	Argentina	Other	1
3	Chile	Maule Valley	1
2	Chile	Colchagua Valley	2
4	France	Alsace	3
5	France	Beaujolais	3
6	Germany	Mosel	1
7	Germany	Rheinhessen	1
9	Italy	Southern Italy	1
8	Italy	Sicily & Sardinia	15
10	Portugal	Douro	1
11	Spain	Northern Spain	2
13	US	Michigan	1
15	US	Virginia	4
14	US	Oregon	5
12	US	California	10

Your turn

If you haven't started the exercise, you can start now.
