Wrangling

May 22, 2018

0.1 Tasks for Step 4

In [1]:
```
# TODO: re-load dataframes from SQL
import csv
import pandas as pd
import numpy as np
import sqlite3
import matplotlib.pyplot as plt
import sklearn.metrics
import matplotlib.pyplot as plt; plt.rcdefaults()
engine=sqlite3.connect('HW1_DB')
airlines_df=pd.read_sql('select* from airlines', engine)
airports_df=pd.read_sql('select* from airports', engine)
routes_df=pd.read_sql('select* from routes', engine)
incidents_df=pd.read_sql('select* from incidents', engine)
```

In [2]:
```
# TODO: airports_df.info(), airlines_df.info(), routes_df.info() and incidents_df.info()
airports_df.info()
airlines_df.info()
routes_df.info()
incidents_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8107 entries, 0 to 8106
Data columns (total 12 columns):
airport_id      8107 non-null int64
airport_name    8107 non-null object
city            8107 non-null object
country         8107 non-null object
iata_ffa        8107 non-null object
icao            8107 non-null object
latitude        8107 non-null float64
longitude       8107 non-null float64
altitude        8107 non-null int64
timezone        8107 non-null float64
dst             8107 non-null object
tz              8107 non-null object
dtypes: float64(3), int64(2), object(7)
```

```
memory usage: 760.1+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6048 entries, 0 to 6047
Data columns (total 8 columns):
airline_id          6048 non-null int64
airline_name        6048 non-null object
airline_alias       6048 non-null object
iata                6048 non-null object
icao                6048 non-null object
airline_callsign    6048 non-null object
country             6048 non-null object
active              6048 non-null object
dtypes: int64(1), object(7)
memory usage: 378.1+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66548 entries, 0 to 66547
Data columns (total 9 columns):
airline_iata_icao           66548 non-null object
airline_id                  66548 non-null int64
source_airport_iata_icao    66548 non-null object
source_airport_id           66548 non-null int64
destination_iata_icao       66548 non-null object
destination_airport_id      66548 non-null int64
codeshare                   66548 non-null object
number_of_stops             66548 non-null int64
Equipments                  66548 non-null object
dtypes: int64(4), object(5)
memory usage: 4.6+ MB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 322 entries, 0 to 321
Data columns (total 3 columns):
Airline      322 non-null object
Date         322 non-null object
FlightNum    322 non-null object
dtypes: object(3)
memory usage: 7.6+ KB
```

0.2 Step 4.0

```
In [3]: #TODO: output airports_df
        airports_df
```

Out[3]:

	airport_id	airport_name	city	\
0	1	Goroka	Goroka	
1	2	Madang	Madang	
2	3	Mount Hagen	Mount Hagen	
3	4	Nadzab	Nadzab	

4	5	Port Moresby Jacksons Intl	Port Moresby
5	6	Wewak Intl	Wewak
6	7	Narsarsuaq	Narssarssuaq
7	8	Nuuk	Godthaab
8	9	Sondre Stromfjord	Sondrestrom
9	10	Thule Air Base	Thule
10	11	Akureyri	Akureyri
11	12	Egilsstadir	Egilsstadir
12	13	Hornafjordur	Hofn
13	14	Husavik	Husavik
14	15	Isafjordur	Isafjordur
15	16	Keflavik International Airport	Keflavik
16	17	Patreksfjordur	Patreksfjordur
17	18	Reykjavik	Reykjavik
18	19	Siglufjordur	Siglufjordur
19	20	Vestmannaeyjar	Vestmannaeyjar
20	21	Sault Ste Marie	Sault Sainte Marie
21	22	Winnipeg St Andrews	Winnipeg
22	23	Shearwater	Halifax
23	24	St Anthony	St. Anthony
24	25	Tofino	Tofino
25	26	Kugaaruk	Pelly Bay
26	27	Baie Comeau	Baie Comeau
27	28	Bagotville	Bagotville
28	29	Baker Lake	Baker Lake
29	30	Campbell River	Campbell River
...	...	...	...
8077	9512	Port of Seattle	Seattle
8078	9513	Ferry Terminal	Whittier
8079	9514	Ferry Terminal	Valdez
8080	9515	Gare Routiere	Saint-Louis
8081	9516	Bus Stop	La Riviere
8082	9517	Bus Stop	Cilaos
8083	9518	Bus Stop	Hell-Bourg
8084	9519	Bus Stop	Salazie
8085	9520	Gare	St Andre
8086	9521	L'Ocean Bus Terminal	Saint Denis
8087	9522	Gare Routiere	Saint Pierre
8088	9523	Baie Ste Anne	Praslin
8089	9524	La Passe	La Digue
8090	9525	Ferry Dock	Bainbridge Island
8091	9526	Bus	Phnom Penh
8092	9527	Bus	Siem Reap
8093	9528	Bus	Sihanoukville
8094	9529	Bus	Kampot
8095	9530	Bus	Kep
8096	9531	Main Station	Taipei
8097	9532	Prominent Hill	Prominent Hill

```
8098        9533                Chatsworth Station          Chatsworth
8099        9534                   Algerciras Port          Algerciras
8100        9535            Ganges Water Aerodrome              Ganges
8101        9536     Pender Harbour Water Aerodrome      Pender Harbour
8102        9537    Mansons Landing Water Aerodrome     Mansons Landing
8103        9538              Port McNeill Airport        Port McNeill
8104        9539       Sullivan Bay Water Aerodrome        Sullivan Bay
8105        9540              Deer Harbor Seaplane         Deer Harbor
8106        9541  San Diego Old Town Transit Center           San Diego

              country iata_ffa  icao   latitude   longitude  altitude  \
0     Papua New Guinea      GKA  AYGA  -6.081689  145.391881      5282
1     Papua New Guinea      MAG  AYMD  -5.207083  145.788700        20
2     Papua New Guinea      HGU  AYMH  -5.826789  144.295861      5388
3     Papua New Guinea      LAE  AYNZ  -6.569828  146.726242       239
4     Papua New Guinea      POM  AYPY  -9.443383  147.220050       146
5     Papua New Guinea      WWK  AYWK  -3.583828  143.669186        19
6            Greenland      UAK  BGBW  61.160517  -45.425978       112
7            Greenland      GOH  BGGH  64.190922  -51.678064       283
8            Greenland      SFJ  BGSF  67.016969  -50.689325       165
9            Greenland      THU  BGTL  76.531203  -68.703161       251
10             Iceland      AEY  BIAR  65.659994  -18.072703         6
11             Iceland      EGS  BIEG  65.283333  -14.401389        76
12             Iceland      HFN  BIHN  64.295556  -15.227222        24
13             Iceland      HZK  BIHU  65.952328  -17.425978        48
14             Iceland      IFJ  BIIS  66.058056  -23.135278         8
15             Iceland      KEF  BIKF  63.985000  -22.605556       171
16             Iceland      PFJ  BIPA  65.555833  -23.965000        11
17             Iceland      RKV  BIRK  64.130000  -21.940556        48
18             Iceland      SIJ  BISI  66.133333  -18.916667        10
19             Iceland      VEY  BIVM  63.424303  -20.278875       326
20              Canada      YAM  CYAM  46.485001  -84.509445       630
21              Canada      YAV  CYAV  50.056389  -97.032500       760
22              Canada      YAW  CYAW  44.639721  -63.499444       167
23              Canada      YAY  CYAY  51.391944  -56.083056       108
24              Canada      YAZ  CYAZ  49.082222 -125.772500        80
25              Canada      YBB  CYBB  68.534444  -89.808056        56
26              Canada      YBC  CYBC  49.132500  -68.204444        71
27              Canada      YBG  CYBG  48.330555  -70.996391       522
28              Canada      YBK  CYBK  64.298889  -96.077778        59
29              Canada      YBL  CYBL  49.950832 -125.270833       346
...                ...      ...   ...        ...         ...       ...
8077     United States           PSEA  47.609178 -122.350605         0
8078     United States           WHIT  60.776566 -148.683367         0
8079     United States           VLDZ  61.123976 -146.365309         0
8080           Reunion           STLS -21.289604   55.407269       103
8081           Reunion           LRIV -21.272574   55.437706       190
8082           Reunion           CILS -21.135518   55.472087      4000
```

8083	Reunion		HLBG	-21.064020	55.518453	3000
8084	Reunion		SLZI	-21.027485	55.538955	3000
8085	Reunion		ANDR	-20.963272	55.652508	100
8086	Reunion		SDNS	-20.877683	55.457393	0
8087	Reunion		STPR	-21.334240	55.471331	0
8088	Seychelles		PLIN	-4.347119	55.765688	0
8089	Seychelles		LADG	-4.347969	55.829069	0
8090	United States		BAIN	47.622237	-122.509362	0
8091	Cambodia		PNMP	11.561716	104.914276	0
8092	Cambodia		SMRP	13.361002	103.859543	0
8093	Cambodia		SNKV	10.607220	103.524886	0
8094	Cambodia		KMPT	10.614922	104.177724	0
8095	Cambodia		KEPC	10.480174	104.294228	0
8096	Taiwan		TPEI	25.046176	121.517532	0
8097	Australia	PXH	YPMH	-29.716667	135.521667	734
8098	United States	CWT		34.256944	-118.598889	978
8099	Spain			36.136000	-5.435000	0
8100	Canada	YGG		48.850000	-123.500000	0
8101	Canada	YPT		49.616667	-124.016667	0
8102	Canada	YMU		50.066667	-124.983333	0
8103	Canada	YMP		50.575556	-127.028611	225
8104	Canada	YTG		50.883333	-126.833333	0
8105	United States	DHB		48.618397	-123.005960	0
8106	United States	OLT		32.755200	-117.199500	0

	timezone	dst	tz
0	10.0	U	Pacific/Port_Moresby
1	10.0	U	Pacific/Port_Moresby
2	10.0	U	Pacific/Port_Moresby
3	10.0	U	Pacific/Port_Moresby
4	10.0	U	Pacific/Port_Moresby
5	10.0	U	Pacific/Port_Moresby
6	-3.0	E	America/Godthab
7	-3.0	E	America/Godthab
8	-3.0	E	America/Godthab
9	-4.0	E	America/Thule
10	0.0	N	Atlantic/Reykjavik
11	0.0	N	Atlantic/Reykjavik
12	0.0	N	Atlantic/Reykjavik
13	0.0	N	Atlantic/Reykjavik
14	0.0	N	Atlantic/Reykjavik
15	0.0	N	Atlantic/Reykjavik
16	0.0	N	Atlantic/Reykjavik
17	0.0	N	Atlantic/Reykjavik
18	0.0	N	Atlantic/Reykjavik
19	0.0	N	Atlantic/Reykjavik
20	-5.0	A	America/Toronto
21	-6.0	A	America/Winnipeg

```
22        -4.0   A       America/Halifax
23        -3.5   A      America/St_Johns
24        -8.0   A     America/Vancouver
25        -7.0   A      America/Edmonton
26        -5.0   A       America/Toronto
27        -5.0   A       America/Toronto
28        -6.0   A      America/Winnipeg
29        -8.0   A     America/Vancouver
...        ...  ..                   ...
8077      -8.0   A   America/Los_Angeles
8078      -9.0   A     America/Anchorage
8079      -9.0   A     America/Anchorage
8080       4.0   E        Indian/Reunion
8081       4.0   E        Indian/Reunion
8082       4.0   E        Indian/Reunion
8083       4.0   E        Indian/Reunion
8084       4.0   E        Indian/Reunion
8085       4.0   E        Indian/Reunion
8086       4.0   E        Indian/Reunion
8087       4.0   E        Indian/Reunion
8088       4.0   N           Indian/Mahe
8089       4.0   N           Indian/Mahe
8090      -8.0   A   America/Los_Angeles
8091       7.0   N       Asia/Phnom_Penh
8092       7.0   N       Asia/Phnom_Penh
8093       7.0   N       Asia/Phnom_Penh
8094       7.0   N       Asia/Phnom_Penh
8095       7.0   N       Asia/Phnom_Penh
8096       8.0   N           Asia/Taipei
8097       9.5   O    Australia/Adelaide
8098      -8.0   A   America/Los_Angeles
8099       1.0   E         Europe/Madrid
8100      -8.0   A     America/Vancouver
8101      -8.0   A     America/Vancouver
8102      -8.0   A     America/Vancouver
8103      -8.0   A     America/Vancouver
8104      -8.0   A     America/Vancouver
8105      -8.0   A   America/Los_Angeles
8106      -8.0   A   America/Los_Angeles

[8107 rows x 12 columns]
```

In [4]:
```
#TODO: output airports_df.describe()
airports_df.describe()
```

Out[4]:

	airport_id	latitude	longitude	altitude	timezone
count	8107.000000	8107.000000	8107.000000	8107.000000	8107.000000
mean	4766.361046	26.817720	-3.921969	933.449365	0.169236

```
std    2943.205193    27.866953    85.900873   1624.740899     5.737326
min       1.000000   -89.999997  -179.877000  -1266.000000   -12.000000
25%    2091.500000     8.824928   -79.022498     38.000000    -5.000000
50%    4257.000000    34.987800     5.292028    272.000000     1.000000
75%    7508.500000    47.957599    49.785821   1020.000000     4.000000
max    9541.000000    82.517778   179.951000  14472.000000    13.000000
```

In [5]:
```
#TODO: output airlines_df
airlines_df
```

Out[5]:

	airline_id	airline_name
0	1	Private flight
1	2	135 Airways
2	3	1Time Airline
3	4	2 Sqn No 1 Elementary Flying Training School
4	5	213 Flight Unit
5	6	223 Flight Unit State Airline
6	7	224th Flight Unit
7	8	247 Jet Ltd
8	9	3D Aviation
9	10	40-Mile Air
10	11	4D Air
11	12	611897 Alberta Limited
12	13	Ansett Australia
13	14	Abacus International
14	15	Abelag Aviation
15	16	Army Air Corps
16	17	Aero Aviation Centre Ltd.
17	18	Aero Servicios Ejecutivos Internacionales
18	19	Aero Biniza
19	20	Aero Albatros
20	21	Aigle Azur
21	22	Aloha Airlines
22	23	Alaska Island Air
23	24	American Airlines
24	25	Aviation Management Corporation
25	26	Atlantis Airlines (USA)
26	27	Aerovista Airlines
27	28	Asiana Airlines
28	29	Askari Aviation
29	30	Australia Asia Airlines
...	...	...
6018	19651	CARICOM AIRWAYS (BARBADOS) INC.
6019	19674	Rainbow Air (RAI)
6020	19675	Rainbow Air Canada
6021	19676	Rainbow Air Polynesia
6022	19677	Rainbow Air Euro
6023	19678	Rainbow Air US

6024	19745	Transilvania
6025	19751	Dobrolet
6026	19774	Spike Airlines
6027	19776	Grand Cru Airlines
6028	19785	Go2Sky
6029	19803	All Argentina
6030	19804	All America
6031	19805	All Asia
6032	19806	All Africa
6033	19807	Regionalia México
6034	19808	All Europe
6035	19809	All Spain
6036	19810	Regional Air Iceland
6037	19811	British Air Ferries
6038	19812	Voestar
6039	19813	All Colombia
6040	19814	Regionalia Uruguay
6041	19815	Regionalia Venezuela
6042	19827	Regionalia Chile
6043	19828	Vuela Cuba
6044	19830	All Australia
6045	19831	Fly Europa
6046	19834	FlyPortugal
6047	19845	FTI Fluggesellschaft

	airline_alias	iata	icao	airline_callsign	country	\
0		-				
1			GNL	GENERAL	United States	
2		1T	RNX	NEXTIME	South Africa	
3			WYT		United Kingdom	
4			TFU		Russia	
5			CHD	CHKALOVSK-AVIA	Russia	
6			TTF	CARGO UNIT	Russia	
7			TWF	CLOUD RUNNER	United Kingdom	
8			SEC	SECUREX	United States	
9		Q5	MLA	MILE-AIR	United States	
10			QRT	QUARTET	Thailand	
11			THD	DONUT	Canada	
12		AN	AAA	ANSETT	Australia	
13		1B			Singapore	
14		W9	AAB	ABG	Belgium	
15			AAC	ARMYAIR	United Kingdom	
16			AAD	SUNRISE	Canada	
17			SII	ASEISA	Mexico	
18			BZS	BINIZA	Mexico	
19			ABM	ALBATROS ESPANA	Spain	
20		ZI	AAF	AIGLE AZUR	France	
21		AQ	AAH	ALOHA	United States	

22			AAK	ALASKA ISLAND	United States
23		AA	AAL	AMERICAN	United States
24			AAM	AM CORP	United States
25			AAO	ATLANTIS AIR	United States
26			AAP	AEROVISTA GROUP	United Arab Emirates
27		OZ	AAR	ASIANA	Republic of Korea
28		4K	AAS	AL-AAS	Pakistan
29			AAU	AUSTASIA	Australia
...	...	...	...	...	...
6018	CARICOM AIRWAYS		CCB		Barbados
6019	Rainbow Air (RAI)	RN	RAB	Rainbow	United States
6020	Rainbow Air CAN	RY	RAY	Rainbow CAN	Canada
6021	Rainbow Air POL	RX	RPO	Rainbow Air	United States
6022	Rainbow Air EU	RU	RUE	Rainbow Air	United Kingdom
6023	Rainbow Air US	RM	RNY	Rainbow Air	United States
6024			TNS		Romania
6025	QD DOB		DOBROLET		Russia
6026	Aero Spike	SO	SAL	Spike Air	United States
6027			GCA		Lithuania
6028			RLX	RELAX	Slovakia
6029	All Argentina	L1	AL1		Argentina
6030	All America	A2	AL2		United States
6031	All Asia	L9	AL3		China
6032	All Africa	9A	99F		South Africa
6033	Regionalia México	N4	J88		Mexico
6034	All Europe	N9	N99		United Kingdom
6035	All Spain	N7	N77		Spain
6036	Regional Air Iceland	9N	N78		Iceland
6037		??	??!		United Kingdom
6038	Voestar Brasil	8K	K88		Brazil
6039	All Colombia	7O	7KK		Colombia
6040	Regionalia Uruguay	2X	2K2		Uruguay
6041	Regionalia Venezuela	9X	9XX		Venezuela
6042	Regionalia Chile	9J	CR1		Chile
6043	Vuela Cuba	6C	6CC		Cuba
6044	All Australia	88	8K8		Australia
6045		ER	RWW		Spain
6046		PO	FPT	FlyPortugal	Portugal
6047			FTI		Germany

	active
0	Y
1	N
2	Y
3	N
4	N
5	N
6	N

7	N
8	N
9	Y
10	N
11	N
12	Y
13	Y
14	N
15	N
16	N
17	N
18	N
19	N
20	Y
21	Y
22	N
23	Y
24	N
25	N
26	N
27	Y
28	Y
29	N
...	...
6018	Y
6019	Y
6020	Y
6021	Y
6022	Y
6023	Y
6024	Y
6025	Y
6026	Y
6027	Y
6028	Y
6029	Y
6030	Y
6031	Y
6032	Y
6033	Y
6034	Y
6035	Y
6036	Y
6037	N
6038	Y
6039	Y
6040	Y
6041	Y

```
6042      Y
6043      Y
6044      Y
6045      Y
6046      Y
6047      N

[6048 rows x 8 columns]
```

In [6]:
```
#TODO: output airlines_df.describe()
airlines_df.describe()
```

Out[6]:
```
         airline_id
count   6048.000000
mean    3844.729167
std     3945.495487
min       -1.000000
25%     1512.750000
50%     3025.500000
75%     4543.250000
max    19845.000000
```

In [7]:
```
#TODO: output routes_df
routes_df
```

Out[7]:
```
      airline_iata_icao  airline_id source_airport_iata_icao  \
0                    2B         410                      AER
1                    2B         410                      ASF
2                    2B         410                      ASF
3                    2B         410                      CEK
4                    2B         410                      CEK
5                    2B         410                      DME
6                    2B         410                      DME
7                    2B         410                      DME
8                    2B         410                      EGO
9                    2B         410                      EGO
10                   2B         410                      GYD
11                   2B         410                      KGD
12                   2B         410                      KZN
13                   2B         410                      KZN
14                   2B         410                      KZN
15                   2B         410                      KZN
16                   2B         410                      KZN
17                   2B         410                      KZN
18                   2B         410                      KZN
19                   2B         410                      LED
20                   2B         410                      LED
21                   2B         410                      LED
22                   2B         410                      MRV
```

```
23                    2B          410                      NBC
24                    2B          410                      NBC
25                    2B          410                      NBC
26                    2B          410                      NBC
27                    2B          410                      NJC
28                    2B          410                      NJC
29                    2B          410                      NUX
...                  ...          ...                      ...
66518                 ZL         4178                      SYD
66519                 ZL         4178                      SYD
66520                 ZL         4178                      SYD
66521                 ZL         4178                      SYD
66522                 ZL         4178                      SYD
66523                 ZL         4178                      SYD
66524                 ZL         4178                      SYD
66525                 ZL         4178                      SYD
66526                 ZL         4178                      SYD
66527                 ZL         4178                      SYD
66528                 ZL         4178                      SYD
66529                 ZL         4178                      SYD
66530                 ZL         4178                      SYD
66531                 ZL         4178                      SYD
66532                 ZL         4178                      SYD
66533                 ZL         4178                      SYD
66534                 ZL         4178                      SYD
66535                 ZL         4178                      TRO
66536                 ZL         4178                      TRO
66537                 ZL         4178                      TSV
66538                 ZL         4178                      TSV
66539                 ZL         4178                      WGA
66540                 ZL         4178                      WGA
66541                 ZL         4178                      WIN
66542                 ZL         4178                      WIN
66543                 ZL         4178                      WYA
66544                 ZM        19016                      DME
66545                 ZM        19016                      FRU
66546                 ZM        19016                      FRU
66547                 ZM        19016                      OSS

       source_airport_id destination_iata_icao  destination_airport_id  \
0                   2965                   KZN                    2990
1                   2966                   KZN                    2990
2                   2966                   MRV                    2962
3                   2968                   KZN                    2990
4                   2968                   OVB                    4078
5                   4029                   KZN                    2990
6                   4029                   NBC                    6969
7                   4029                   UUA                    6160
```

8	6156	KGD	2952
9	6156	KZN	2990
10	2922	NBC	6969
11	2952	EGO	6156
12	2990	AER	2965
13	2990	ASF	2966
14	2990	CEK	2968
15	2990	DME	4029
16	2990	EGO	6156
17	2990	LED	2948
18	2990	SVX	2975
19	2948	KZN	2990
20	2948	NBC	6969
21	2948	UUA	6160
22	2962	ASF	2966
23	6969	DME	4029
24	6969	GYD	2922
25	6969	LED	2948
26	6969	SVX	2975
27	2972	SVX	2975
28	2972	UUA	6160
29	4364	SVX	2975
...	...	...	...
66518	3361	ABX	3333
66519	3361	ARM	4062
66520	3361	BHQ	6241
66521	3361	BHS	6249
66522	3361	BNK	4291
66523	3361	DBO	3358
66524	3361	GFF	6271
66525	3361	LSY	6287
66526	3361	MIM	6294
66527	3361	MQL	6298
66528	3361	MYA	6302
66529	3361	NRA	6307
66530	3361	NTL	4320
66531	3361	OAG	6793
66532	3361	PKE	6317
66533	3361	TRO	6794
66534	3361	WGA	3363
66535	6794	GFN	6792
66536	6794	SYD	3361
66537	3330	ISA	3324
66538	3330	WIN	6337
66539	3363	MEL	3339
66540	3363	SYD	3361
66541	6337	LRE	6289
66542	6337	TSV	3330

66543	6334	ADL	3341
66544	4029	FRU	2912
66545	2912	DME	4029
66546	2912	OSS	2913
66547	2913	FRU	2912

	codeshare	number_of_stops	Equipments
0		0	CR2
1		0	CR2
2		0	CR2
3		0	CR2
4		0	CR2
5		0	CR2
6		0	CR2
7		0	CR2
8		0	CR2
9		0	CR2
10		0	CR2
11		0	CR2
12		0	CR2
13		0	CR2
14		0	CR2
15		0	CR2
16		0	CR2
17		0	CR2
18		0	CR2
19		0	CR2
20		0	CR2
21		0	CR2
22		0	CR2
23		0	CR2
24		0	CR2
25		0	CR2
26		0	CR2
27		0	CR2
28		0	CR2
29		0	CR2
...	...	...	...
66518		0	SF3
66519		0	SF3
66520		0	SF3
66521		0	SF3
66522		0	SF3
66523		0	SF3
66524		0	SF3
66525		0	SF3
66526		0	SF3
66527		0	SF3

```
66528                           0        SF3
66529                           0        SF3
66530                           0        SF3
66531                           0        SF3
66532                           0        SF3
66533                           0        SF3
66534                           0        SF3
66535                           0        SF3
66536                           0        SF3
66537                           0        SF3
66538                           0        SF3
66539                           0        SF3
66540                           0        SF3
66541                           0        SF3
66542                           0        SF3
66543                           0        SF3
66544                           0        734
66545                           0        734
66546                           0        734
66547                           0        734

[66548 rows x 9 columns]
```

In [8]:
```
#TODO: output routes_df.describe()
routes_df.describe()
```

Out[8]:
```
          airline_id  source_airport_id  destination_airport_id  \
count  66548.000000       66548.000000            66548.000000
mean    3367.569754        2660.201929             2662.088192
std     3182.024330        1577.965637             1578.661433
min       10.000000           1.000000                1.000000
25%     1317.000000        1382.000000             1382.000000
50%     2942.000000        2958.000000             2958.000000
75%     4608.000000        3670.000000             3670.000000
max    19676.000000        9327.000000             9327.000000

       number_of_stops
count     66548.000000
mean          0.000165
std           0.012856
min           0.000000
25%           0.000000
50%           0.000000
75%           0.000000
max           1.000000
```

In [9]:
```
#TODO: output incidents_df
incidents_df
```

Out[9]:

	Airline	Date	FlightNum
0	Comair	1997-01-09 00:00:00	3272
1	Stavropolskaya Aktsionernaya Avia	1997-03-18 00:00:00	1023
2	Merpati Nusantara Airlines	1997-04-19 00:00:00	106
3	China Southern Airlines	1997-05-08 00:00:00	3456
4	FedEx Express	1997-07-31 00:00:00	14
5	Sempati Air	1997-07-17 00:00:00	304
6	Korean Air	1997-08-06 00:00:00	801
7	Formosa Airlines	1997-08-10 00:00:00	7601
8	Vietnam Airlines	1997-09-03 00:00:00	815
9	Royal Brunei Airlines	1997-09-06 00:00:00	238
10	Garuda Indonesia	1997-09-26 00:00:00	152
11	Austral Líneas Aéreas	1997-10-10 00:00:00	2553
12	Tajikistan Airlines	1997-12-15 00:00:00	3183
13	Aerosvit	1997-12-17 00:00:00	241
14	SilkAir	1997-12-19 00:00:00	185
15	United Airlines	1997-12-28 00:00:00	826
16	Cebu Pacific	1998-02-02 00:00:00	387
17	China Airlines	1998-02-16 00:00:00	676
18	Philippine Airlines	1998-03-22 00:00:00	137
19	Air France	1998-04-20 00:00:00	422
20	PIA	1998-05-25 00:00:00	544
21	Propair	1998-06-18 00:00:00	420
22	Proteus Airlines	1998-07-30 00:00:00	706:
23	Myanma Airways	1998-08-24 00:00:00	635
24	Cubana de Aviación	1998-08-29 00:00:00	389
25	Swissair	1998-09-02 00:00:00	111
26	PauknAir	1998-09-25 00:00:00	4101
27	Lionair	1998-09-29 00:00:00	602
28	Thai Airways International	1998-12-11 00:00:00	261
29	China Southwest Airlines	1999-02-24 00:00:00	4509
..	...	...	...
292	Ceiba Intercontinental Airlines	2015-09-05 00:00:00	71
293	British Airways	2015-09-08 00:00:00	2276
294	Aviastar	2015-10-02 00:00:00	7503
295	Dynamic Airways	2015-10-29 00:00:00	405
296	Metrojet	2015-10-31 00:00:00	9268
297	West Air Sweden	2016-01-08 00:00:00	294
298	Daallo Airlines	2016-02-02 00:00:00	159
299	Tara Air	2016-02-24 00:00:00	193
300	Flydubai	2016-03-19 00:00:00	981
301	EgyptAir	2016-03-29 00:00:00	181
302	Batik Air	2016-04-04 00:00:00	7703
303	EgyptAir	2016-05-19 00:00:00	804
304	Korean Air	2016-05-27 00:00:00	2708
305	Singapore Airlines	2016-06-27 00:00:00	368
306	Emirates	2016-08-03 00:00:00	521
307	ASL Airlines Hungary	2016-08-05 00:00:00	7332

```
308                 Southwest Airlines  2016-08-27 00:00:00      3472
309                  American Airlines  2016-10-28 00:00:00       383
310                      FedEx Express  2016-10-28 00:00:00       910
311                       LaMia Airlines  2016-11-28 00:00:00      2933
312    Pakistan International Airlines  2016-12-07 00:00:00       661
313                            Aerosucre  2016-12-20 00:00:00      4544
314                   Afriqiyah Airways  2016-12-23 00:00:00       209
315                    Turkish Airlines  2017-01-16 00:00:00      6491
316                  Peruvian Airlines  2017-03-28 00:00:00       112
317                          Summit Air  2017-05-27 00:00:00       409
318                          Air Canada  2017-07-07 00:00:00       759
319                          Air France  2017-09-30 00:00:00        66
320                 West Wind Aviation  2017-12-13 00:00:00       280
321                    Pegasus Airlines  2018-01-13 00:00:00      8622

[322 rows x 3 columns]
```

```
In [10]: #TODO: output incidents_df.describe()
         incidents_df.describe()
```

Out[10]:

	Airline	Date	FlightNum
count	322	322	322
unique	240	309	301
top	Turkish Airlines	2001-09-11 00:00:00	103
freq	6	4	3

0.3 Step 4.1

```
In [11]: # TODO: output dictionary {airports: ___, destinations: ___, sources: ___}
         s_id=routes_df.source_airport_id.unique()
         d_id=routes_df.destination_airport_id.unique()
         no_of_source_airports=len(s_id)
         no_of_dest_airports=len(d_id)
         total_airports_ids=len(np.unique(airports_df.airport_id))
         data_dict={'airports':total_airports_ids,'destinations':no_of_dest_airports,'sources':n
         data_dict
```

```
Out[11]: {'airports': 8107, 'destinations': 3170, 'sources': 3167}
```

0.4 Step 4.2.1

```
In [12]: # TODO: output airports you can fly to by not from
         x=(set(routes_df.destination_airport_id.unique()).difference(set(routes_df.source_airpo
         x

         list1=[]
         for i in x:
             list1.append(i)
         temp_dict={'temp_col':list1}
```

```
dict_df=pd.DataFrame.from_dict(temp_dict)
name_airports=pd.merge(airports_df,dict_df, left_on='airport_id', right_on='temp_col',h
name_airports
```

Out[12]:
```
0                            Filippos
1                  Coronel E Carvajal
2     Teniente Coronel Luis A Mantilla
3                         Utila Airport
4              Abraham Lincoln Capital
5                  Punta Gorda Airport
6                  Dalanzadgad Airport
7                  Pilot Point Airport
8                      Karuluk Airport
9          Port Williams Seaplane Base
10            Zachar Bay Seaplane Base
11                Campo Alegre Airport
12                      Breves Airport
13                         El Porvenir
14                  Oriximina Airport
Name: airport_name, dtype: object
```

0.5 Step 4.2.2

In [13]:
```
#TODO: top 10 most popular airports
name_airports=pd.merge(airports_df,routes_df, left_on='airport_id', right_on='destinati

df_top_freq = name_airports.groupby(['airport_id', 'airport_name'])['destination_airpor
    {"count": len}).sort_values(
    "count", ascending=False).head(10).reset_index()
df_top_freq
```

Out[13]:

	airport_id	airport_name	count
0	3682	Hartsfield Jackson Atlanta Intl	905
1	3830	Chicago Ohare Intl	550
2	3364	Capital Intl	526
3	507	Heathrow	524
4	1382	Charles De Gaulle	517
5	3484	Los Angeles Intl	497
6	340	Frankfurt Main	493
7	3670	Dallas Fort Worth Intl	467
8	3797	John F Kennedy Intl	455
9	580	Schiphol	450

0.6 Step 4.3

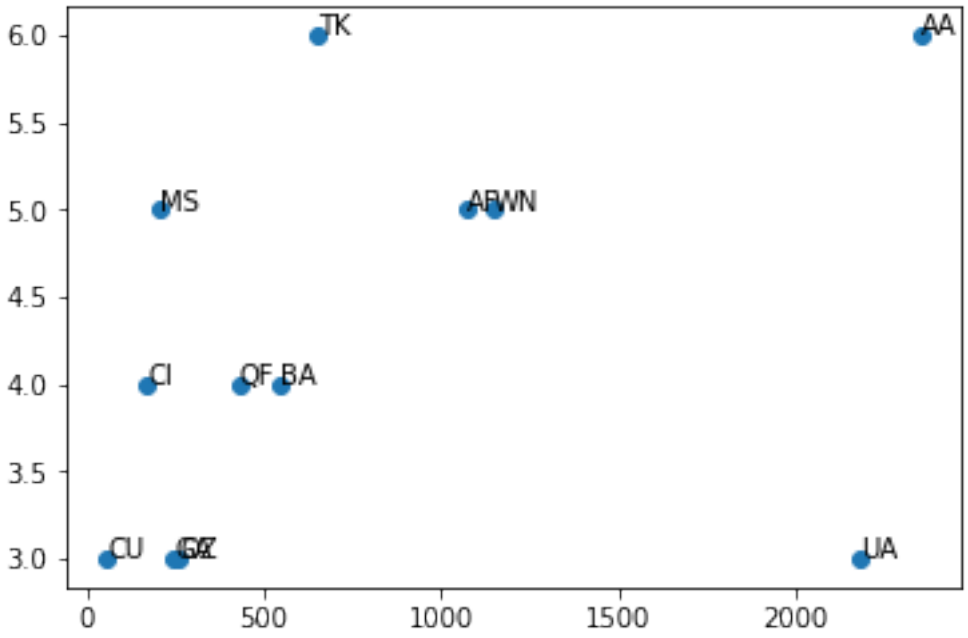
In [16]:
```
# TODO: Create dataframe of airline_iata, number of incidents, number of routes
data_frame=incidents_df
data_frame['Airline']=data_frame['Airline'].str.lower()
```

```
df1 = data_frame.groupby(['Airline'])['Airline'].agg(
    {"incident_count": len}).sort_values(
    "incident_count", ascending=False).reset_index()


data_frame1=airlines_df
data_frame1['airline_name']=data_frame1['airline_name'].str.lower()
merged_frame=pd.merge(df1,data_frame1, left_on='Airline', right_on='airline_name')

df2 = routes_df.groupby(['airline_id'])['airline_id'].agg(
    {"route_count": len}).sort_values(
    "route_count", ascending=False).reset_index()
#print(merged_frame.columns)
#print(df2.columns)
df3=pd.merge(merged_frame,df2, left_on='airline_id', right_on='airline_id')

iff=df3[['incident_count', 'iata', 'route_count']]

iff=iff[iff.incident_count > 2]
iff
```

Out[16]:

	incident_count	iata	route_count
0	6	AA	2354
1	6	TK	652
2	5	MS	205
3	5	AF	1071
4	5	WN	1146
5	4	QF	430
6	4	CI	172
7	4	BA	549
8	3	OZ	262
9	3	GA	248
10	3	UA	2180
11	3	CU	59

In [17]:
```
# TODO: Scatter plot of number of routes vs. number of incidents
routes=list(iff.route_count)
no_of_incidents=list(iff.incident_count)
iata=list(iff.iata)

fig, ax = plt.subplots()
ax.scatter(routes,no_of_incidents)
for i, txt in enumerate(iata):
    ax.annotate(txt, (routes[i],no_of_incidents[i]))
```



0.7 Tasks for Step 5

```
In [18]: # TODO: get random sample of flights from SQL
         flights_s_df=pd.read_sql_query('SELECT * FROM flights ORDER BY RANDOM() LIMIT 1000', co
         flights_s_df
```

Out[18]:

	Year	Month	DayofMonth	AirlineID	Carrier	FlightNum	Origin	Dest	\
0	2016	10	30	19393	WN	3071	SAT	PHX	
1	2016	10	7	19790	DL	839	SRQ	ATL	
2	2016	10	26	19805	AA	768	MCO	PHL	
3	2016	10	27	19977	UA	1075	LAS	EWR	
4	2016	10	4	19393	WN	2285	MCO	BUF	
5	2016	10	2	20409	B6	1551	ACK	BOS	
6	2016	10	25	19805	AA	320	SAT	DFW	
7	2016	10	14	20304	OO	5899	RNO	SFO	
8	2016	10	15	20304	OO	5254	TUS	ORD	
9	2016	10	23	21171	VX	251	JFK	LAS	
10	2016	10	5	19790	DL	2371	MSP	SMF	
11	2016	10	10	20366	EV	5528	TYS	ATL	
12	2016	10	16	19393	WN	1103	BNA	MSY	
13	2016	10	5	19393	WN	2852	SAN	SMF	
14	2016	10	8	19393	WN	1007	AUS	DEN	
15	2016	10	22	19393	WN	3314	BUR	OAK	
16	2016	10	13	19977	UA	663	EWR	IAH	
17	2016	10	2	19805	AA	1936	CLT	EWR	
18	2016	10	28	19393	WN	951	MSY	BWI	

19	2016	10	7	19393	WN	1840	RNO	LAX
20	2016	10	27	19977	UA	380	SFO	BOS
21	2016	10	23	19805	AA	350	LAX	SFO
22	2016	10	10	20304	OO	5020	LAX	SBP
23	2016	10	19	19790	DL	585	DTW	LAX
24	2016	10	24	19393	WN	259	OKC	HOU
25	2016	10	6	19393	WN	2884	DAL	ABQ
26	2016	10	2	19790	DL	1726	ATL	FLL
27	2016	10	27	19977	UA	1659	CVG	ORD
28	2016	10	20	19393	WN	1291	MDW	GRR
29	2016	10	16	19977	UA	640	ORD	PHL
..	...	...	...	...	...	...	...	...
970	2016	10	6	20304	OO	4585	MSN	SLC
971	2016	10	21	19393	WN	3653	LAX	SMF
972	2016	10	13	19790	DL	1386	ATL	DSM
973	2016	10	25	19393	WN	2006	LAS	SFO
974	2016	10	23	19393	WN	2082	OMA	STL
975	2016	10	21	20409	B6	260	PHL	BOS
976	2016	10	3	19790	DL	630	DTW	ATL
977	2016	10	30	20366	EV	5263	RIC	BOS
978	2016	10	2	19790	DL	1739	ATL	TUS
979	2016	10	9	19393	WN	3146	PDX	MDW
980	2016	10	16	19790	DL	1903	ATL	FAY
981	2016	10	17	20304	OO	4752	SLC	EUG
982	2016	10	19	19393	WN	2357	OAK	SEA
983	2016	10	18	20304	OO	4759	SLC	PSC
984	2016	10	7	19790	DL	2478	ATL	JAN
985	2016	10	21	19790	DL	1363	GRR	DTW
986	2016	10	22	19805	AA	2278	LAX	ORD
987	2016	10	20	20304	OO	5151	SGU	DEN
988	2016	10	19	20409	B6	1507	JFK	IAD
989	2016	10	5	19393	WN	1668	SAN	SMF
990	2016	10	7	19805	AA	1205	SNA	ORD
991	2016	10	30	20366	EV	2845	ACT	DFW
992	2016	10	14	19790	DL	1307	RIC	ATL
993	2016	10	11	20366	EV	4430	EWR	SAV
994	2016	10	24	20409	B6	24	LAX	JFK
995	2016	10	15	19977	UA	261	SFO	LAX
996	2016	10	10	20366	EV	5264	CRW	ATL
997	2016	10	3	19393	WN	884	MDW	MSP
998	2016	10	22	20304	OO	4747	CID	DTW
999	2016	10	22	20436	F9	1637	MCO	MKE

	ArrDelayMinutes	Cancelled
0	41.0	0.0
1	0.0	0.0
2	71.0	0.0
3	134.0	0.0

4	0.0	0.0
5	0.0	0.0
6	0.0	0.0
7	NaN	1.0
8	0.0	0.0
9	0.0	0.0
10	0.0	0.0
11	0.0	0.0
12	7.0	0.0
13	6.0	0.0
14	0.0	0.0
15	0.0	0.0
16	0.0	0.0
17	0.0	0.0
18	0.0	0.0
19	5.0	0.0
20	0.0	0.0
21	0.0	0.0
22	0.0	0.0
23	0.0	0.0
24	0.0	0.0
25	0.0	0.0
26	0.0	0.0
27	0.0	0.0
28	0.0	0.0
29	0.0	0.0
..	...	...
970	0.0	0.0
971	0.0	0.0
972	0.0	0.0
973	0.0	0.0
974	0.0	0.0
975	0.0	0.0
976	0.0	0.0
977	45.0	0.0
978	0.0	0.0
979	0.0	0.0
980	0.0	0.0
981	0.0	0.0
982	0.0	0.0
983	0.0	0.0
984	0.0	0.0
985	19.0	0.0
986	0.0	0.0
987	1.0	0.0
988	0.0	0.0
989	0.0	0.0
990	0.0	0.0

```
991              0.0        0.0
992              0.0        0.0
993              0.0        0.0
994              0.0        0.0
995              0.0        0.0
996              0.0        0.0
997             48.0        0.0
998              0.0        0.0
999              5.0        0.0

[1000 rows x 10 columns]
```

In [19]:
```
# TODO: create nested map of Jaccard distance
#df_r=routes_df.drop_duplicates(subset=None, keep='first', inplace=False)
#df_f=flights_s_df.drop_duplicates(subset=None, keep='first', inplace=False)


dict2={}
for element in routes_df:
    element_rcol=routes_df[element].drop_duplicates()
    dict1={}
    my_list=[]
    for item in flights_s_df:
        if np.issubdtype(flights_s_df[item].dtype,np.number):
            continue
        else:
            item_fcol=flights_s_df[item].drop_duplicates()
            intersection=pd.Series(list(set(element_rcol).intersection(set(item_fcol))))
            union=pd.Series(list(set(element_rcol).union(set(item_fcol))))
        if len(intersection)/len(union)==0:
            continue
        else:
            my_list.append(len(intersection)/len(union))
            dict1.update({item:my_list})
            dict2.update({element:dict1})


jaccard_map=dict2
```

0.8 Step 5.1

In [20]:
```
# TODO: output nested map
jaccard_map
```

Out[20]:
```
{'airline_iata_icao': {'Carrier': [0.01845018450184502]},
 'destination_iata_icao': {'Dest': [0.04604225796278776, 0.04290220820189274],
  'Origin': [0.04604225796278776, 0.04290220820189274]},
 'source_airport_iata_icao': {'Dest': [0.045755758914484065,
```

```
 0.0429428481212504],
'Origin': [0.045755758914484065, 0.0429428481212504]}}
```

0.9 Step 5.2.1

In [21]:
```
# TODO: create joint DataFrame flights_s_df
flights_s_df=routes_df.merge(flights_s_df, left_on=['airline_iata_icao','destination_ia
flights_s_df
```

Out[21]:

	airline_iata_icao	airline_id	source_airport_iata_icao	source_airport_id	\
0	AA	24	ATL	3682	
1	AA	24	ATL	3682	
2	AA	24	AUS	3673	
3	AA	24	BOS	3448	
4	AA	24	BWI	3849	
5	AA	24	BWI	3849	
6	AA	24	BWI	3849	
7	AA	24	BWI	3849	
8	AA	24	CLT	3876	
9	AA	24	CLT	3876	
10	AA	24	CLT	3876	
11	AA	24	CLT	3876	
12	AA	24	CLT	3876	
13	AA	24	CLT	3876	
14	AA	24	CLT	3876	
15	AA	24	CLT	3876	
16	AA	24	CLT	3876	
17	AA	24	CLT	3876	
18	AA	24	CLT	3876	
19	AA	24	DCA	3520	
20	AA	24	DCA	3520	
21	AA	24	DCA	3520	
22	AA	24	DEN	3751	
23	AA	24	DFW	3670	
24	AA	24	DFW	3670	
25	AA	24	DFW	3670	
26	AA	24	DFW	3670	
27	AA	24	DFW	3670	
28	AA	24	DFW	3670	
29	AA	24	DFW	3670	
..	...	...	...	...	
734	WN	4547	SJC	3748	
735	WN	4547	SJC	3748	
736	WN	4547	SJC	3748	
737	WN	4547	SJC	3748	
738	WN	4547	SLC	3536	
739	WN	4547	SLC	3536	
740	WN	4547	SLC	3536	

```
741                WN        4547                     SLC                3536
742                WN        4547                     SLC                3536
743                WN        4547                     SMF                3817
744                WN        4547                     SMF                3817
745                WN        4547                     SMF                3817
746                WN        4547                     SMF                3817
747                WN        4547                     SMF                3817
748                WN        4547                     SMF                3817
749                WN        4547                     SNA                3867
750                WN        4547                     STL                3678
751                WN        4547                     STL                3678
752                WN        4547                     STL                3678
753                WN        4547                     STL                3678
754                WN        4547                     STL                3678
755                WN        4547                     STL                3678
756                WN        4547                     TPA                3646
757                WN        4547                     TPA                3646
758                WN        4547                     TPA                3646
759                WN        4547                     TPA                3646
760                WN        4547                     TUL                3855
761                WN        4547                     TUS                3636
762                WN        4547                     TUS                3636
763                WN        4547                     TUS                3636

    destination_iata_icao  destination_airport_id codeshare  number_of_stops  \
0                     CLT                   3876         Y                0
1                     DFW                   3670                          0
2                     DFW                   3670                          0
3                     ORD                   3830                          0
4                     CLT                   3876                          0
5                     CLT                   3876                          0
6                     CLT                   3876                          0
7                     CLT                   3876                          0
8                     BDL                   3825                          0
9                     BNA                   3690                          0
10                    BNA                   3690                          0
11                    BWI                   3849                          0
12                    DTW                   3645                          0
13                    DTW                   3645                          0
14                    JFK                   3797                          0
15                    LAS                   3877                          0
16                    LAX                   3484                          0
17                    MCO                   3878                          0
18                    TPA                   3646                          0
19                    DFW                   3670                          0
20                    MIA                   3576                          0
21                    MIA                   3576                          0
22                    PHX                   3462                          0
```

23	BDL	3825		0
24	CMH	3759		0
25	COS	3819		0
26	HNL	3728		0
27	LAX	3484		0
28	MCI	3458		0
29	MIA	3576		0
..	...	...	...	...
734	BUR	3644		0
735	BUR	3644		0
736	LAS	3877		0
737	LAX	3484		0
738	DEN	3751		0
739	LAS	3877		0
740	MDW	3747		0
741	OAK	3453		0
742	OAK	3453		0
743	BUR	3644		0
744	LAS	3877		0
745	LAX	3484		0
746	ONT	3734		0
747	SAN	3731		0
748	SAN	3731		0
749	SFO	3469		0
750	BOS	3448		0
751	FLL	3533		0
752	LAS	3877		0
753	OMA	3454		0
754	PHX	3462		0
755	TPA	3646		0
756	BNA	3690		0
757	LAS	3877		0
758	MDW	3747		0
759	MSY	3861		0
760	PHX	3462		0
761	LAS	3877		0
762	LAX	3484		0
763	LAX	3484		0

	Equipments	Year	Month	DayofMonth	AirlineID	Carrier	\
0	319 320 CR9 321	2016	10	30	19805	AA	
1	M80 319 M83	2016	10	22	19805	AA	
2	M80 738 M83	2016	10	23	19805	AA	
3	738	2016	10	19	19805	AA	
4	E90 321 757 319 737 320	2016	10	7	19805	AA	
5	E90 321 757 319 737 320	2016	10	22	19805	AA	
6	E90 321 757 319 737 320	2016	10	27	19805	AA	
7	E90 321 757 319 737 320	2016	10	24	19805	AA	

8			319	321	320	2016	10	4	19805	AA
9				319	CR7	2016	10	6	19805	AA
10				319	CR7	2016	10	26	19805	AA
11	319	737	321	757	320	E90	2016	10	26	19805	AA
12			319	321	320	2016	10	3	19805	AA
13			319	321	320	2016	10	6	19805	AA
14				321	319	2016	10	29	19805	AA
15					321	2016	10	4	19805	AA
16					321	2016	10	18	19805	AA
17		321	767	757	319	2016	10	16	19805	AA
18		320	321	319	737	2016	10	2	19805	AA
19				738	M80	2016	10	26	19805	AA
20					738	2016	10	8	19805	AA
21					738	2016	10	25	19805	AA
22				320	321	2016	10	27	19805	AA
23					738	2016	10	26	19805	AA
24				M83	M80	2016	10	12	19805	AA
25				M80	M83	2016	10	26	19805	AA
26					763	2016	10	21	19805	AA
27		738	763	757	M83	2016	10	17	19805	AA
28				M80	M83	2016	10	22	19805	AA
29			738	763	757	2016	10	23	19805	AA
..					...	...	...	...	...	...
734					73W	2016	10	3	19393	WN
735					73W	2016	10	22	19393	WN
736			73W	733	73C	73H	2016	10	9	19393	WN
737		73W	733	73C	73H	738	2016	10	4	19393	WN
738		73W	738	733	73C	73H	2016	10	30	19393	WN
739				73W	73C	2016	10	14	19393	WN
740				73H	73W	2016	10	13	19393	WN
741			73W	73C	733	2016	10	10	19393	WN
742			73W	73C	733	2016	10	13	19393	WN
743					73W	2016	10	13	19393	WN
744		73H	73W	733	73C	738	2016	10	3	19393	WN
745		73C	73W	733	73H	738	2016	10	21	19393	WN
746			733	73W	73C	73H	2016	10	16	19393	WN
747			73W	73H	73C	733	2016	10	5	19393	WN
748			73W	73H	73C	733	2016	10	5	19393	WN
749					73W	2016	10	3	19393	WN
750				73C	73W	2016	10	3	19393	WN
751				733	73W	2016	10	20	19393	WN
752		738	73W	73C	73H	733	2016	10	15	19393	WN
753			733	73C	73W	2016	10	23	19393	WN
754			73W	733	73H	2016	10	28	19393	WN
755			73W	73C	73H	2016	10	24	19393	WN
756			73C	73W	733	2016	10	1	19393	WN
757				73H	73W	2016	10	10	19393	WN
758				73H	73W	2016	10	4	19393	WN

759		73W 733	2016	10	31	19393	WN
760		73W	2016	10	12	19393	WN
761		73W 73C	2016	10	30	19393	WN
762	73C 73W 733	73H	2016	10	2	19393	WN
763	73C 73W 733	73H	2016	10	18	19393	WN

	FlightNum	Origin	Dest	ArrDelayMinutes	Cancelled
0	589	CLT	ATL	10.0	0.0
1	1473	DFW	ATL	0.0	0.0
2	1561	DFW	AUS	0.0	0.0
3	1205	ORD	BOS	0.0	0.0
4	1998	CLT	BWI	0.0	0.0
5	703	CLT	BWI	0.0	0.0
6	1924	CLT	BWI	0.0	0.0
7	1960	CLT	BWI	0.0	0.0
8	1837	BDL	CLT	0.0	0.0
9	2065	BNA	CLT	3.0	0.0
10	2065	BNA	CLT	0.0	0.0
11	1798	BWI	CLT	7.0	0.0
12	756	DTW	CLT	3.0	0.0
13	1713	DTW	CLT	0.0	0.0
14	148	JFK	CLT	0.0	0.0
15	431	LAS	CLT	33.0	0.0
16	696	LAX	CLT	0.0	0.0
17	746	MCO	CLT	0.0	0.0
18	1719	TPA	CLT	0.0	0.0
19	2290	DFW	DCA	0.0	0.0
20	2239	MIA	DCA	7.0	0.0
21	1768	MIA	DCA	0.0	0.0
22	1704	PHX	DEN	12.0	0.0
23	1581	BDL	DFW	0.0	0.0
24	938	CMH	DFW	0.0	0.0
25	160	COS	DFW	0.0	0.0
26	8	HNL	DFW	0.0	0.0
27	2473	LAX	DFW	11.0	0.0
28	1046	MCI	DFW	0.0	0.0
29	1541	MIA	DFW	0.0	0.0
..	...	...	...	...	...
734	2542	BUR	SJC	6.0	0.0
735	610	BUR	SJC	0.0	0.0
736	562	LAS	SJC	10.0	0.0
737	1796	LAX	SJC	20.0	0.0
738	3125	DEN	SLC	0.0	0.0
739	416	LAS	SLC	11.0	0.0
740	1078	MDW	SLC	0.0	0.0
741	2933	OAK	SLC	0.0	0.0
742	2933	OAK	SLC	0.0	0.0
743	1754	BUR	SMF	0.0	0.0

```
744        232    LAS  SMF              0.0        0.0
745       3653    LAX  SMF              0.0        0.0
746        747    ONT  SMF             39.0        0.0
747       2852    SAN  SMF              6.0        0.0
748       1668    SAN  SMF              0.0        0.0
749       2212    SFO  SNA              0.0        0.0
750       3099    BOS  STL             11.0        0.0
751        826    FLL  STL              0.0        0.0
752       3499    LAS  STL              0.0        0.0
753       2082    OMA  STL              0.0        0.0
754       1924    PHX  STL             23.0        0.0
755        499    TPA  STL              0.0        0.0
756       1209    BNA  TPA              0.0        0.0
757       2955    LAS  TPA              0.0        0.0
758        806    MDW  TPA              1.0        0.0
759       1487    MSY  TPA             33.0        0.0
760       2570    PHX  TUL             55.0        0.0
761        556    LAS  TUS             25.0        0.0
762       2436    LAX  TUS              0.0        0.0
763       2436    LAX  TUS             11.0        0.0

[764 rows x 19 columns]
```

In [24]:
```
# TODO: output for each airline IATA code and destination city IATA code, the number of
fre_df=flights_s_df.groupby(['airline_iata_icao','destination_iata_icao'])['airline_iat
    {"count": len}).sort_values(
    "count", ascending=False).reset_index()
fre_df
```

Out[24]:

	airline_iata_icao	destination_iata_icao	count
0	DL	ATL	56
1	UA	ORD	20
2	AA	DFW	19
3	AA	CLT	16
4	WN	LAS	15
5	WN	LAX	13
6	WN	MDW	13
7	B6	BOS	13
8	AA	MIA	13
9	DL	MSP	13
10	WN	DEN	12
11	B6	JFK	12
12	DL	DTW	12
13	UA	EWR	12
14	AA	PHX	11
15	AS	SEA	11
16	UA	DEN	10
17	WN	BWI	10

18	UA	IAH	10
19	WN	MCO	9
20	DL	LAX	8
21	WN	SJC	8
22	DL	SLC	8
23	WN	SAN	8
24	UA	SFO	7
25	WN	OAK	7
26	AA	ORD	7
27	WN	TPA	6
28	UA	LAX	6
29	WN	PHX	6
..	...	...	...
214	B6	PBI	1
215	UA	PHL	1
216	B6	BDL	1
217	B6	BUF	1
218	UA	OMA	1
219	UA	MSY	1
220	B6	DCA	1
221	B6	DEN	1
222	B6	DFW	1
223	B6	EWR	1
224	B6	HPN	1
225	UA	DTW	1
226	B6	LAX	1
227	B6	OAK	1
228	UA	BNA	1
229	HA	LAX	1
230	UA	ABQ	1
231	NK	RSW	1
232	NK	PHL	1
233	B6	PDX	1
234	NK	OAK	1
235	NK	MYR	1
236	B6	PSE	1
237	NK	LAS	1
238	B6	RIC	1
239	B6	ROC	1
240	NK	DEN	1
241	HA	OGG	1
242	HA	OAK	1
243	DL	RIC	1

[244 rows x 3 columns]

0.10 Step 5.2.2

In [32]:
```
# TODO: output bad flights
df_1=flights_s_df[(flights_s_df.ArrDelayMinutes >= 30)]
df_2=flights_s_df[flights_s_df.Cancelled > 0 ]
df1=df_1.append(df_2)
fre_df1=df1.groupby(['airline_iata_icao','destination_iata_icao'])['airline_iata_icao']
    {"count": len}).sort_values(
    "count", ascending=False).reset_index()
fre_df1
```

Out[32]:

	airline_iata_icao	destination_iata_icao	count
0	B6	BOS	5
1	WN	LAX	4
2	UA	ORD	3
3	AA	MIA	3
4	DL	ATL	3
5	WN	MDW	2
6	UA	BOS	2
7	AA	DFW	2
8	F9	DEN	2
9	B6	JFK	2
10	UA	EWR	2
11	AA	CLT	2
12	AA	PHX	2
13	WN	SAT	2
14	WN	MSY	1
15	WN	ONT	1
16	WN	PDX	1
17	WN	DEN	1
18	WN	BOS	1
19	VX	SFO	1
20	VX	IAD	1
21	WN	PHX	1
22	UA	SFO	1
23	WN	PIT	1
24	UA	LAS	1
25	HA	LAX	1
26	UA	BWI	1
27	NK	LGA	1
28	F9	MCO	1
29	DL	SLC	1
30	DL	LAX	1
31	DL	IAD	1
32	DL	CLT	1
33	B6	HPN	1
34	B6	FLL	1
35	B6	DCA	1
36	AS	SEA	1

37	AS	EWR	1
38	AA	SAT	1
39	AA	MCO	1
40	AA	LAS	1
41	WN	SNA	1

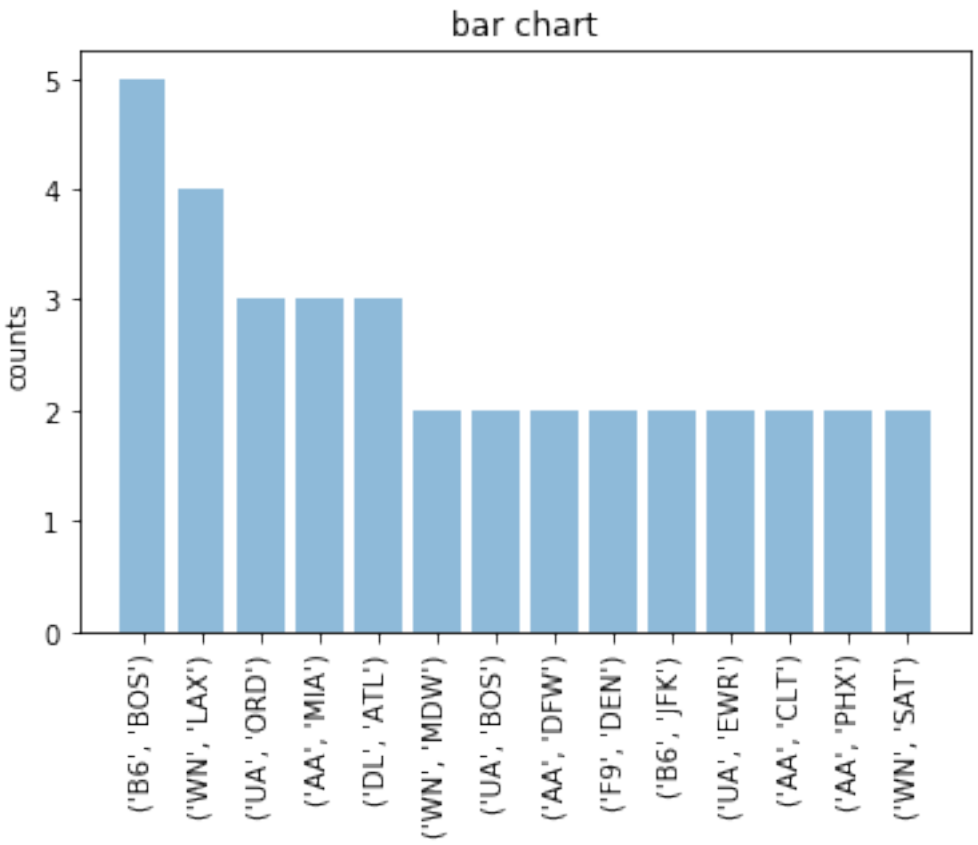
0.11 Step 5.2.3

In [36]:
```
# TODO: pairs with more than one bad flight
freq=fre_df1[(fre_df1['count'])>1]
freq
```

Out[36]:

	airline_iata_icao	destination_iata_icao	count
0	B6	BOS	5
1	WN	LAX	4
2	UA	ORD	3
3	AA	MIA	3
4	DL	ATL	3
5	WN	MDW	2
6	UA	BOS	2
7	AA	DFW	2
8	F9	DEN	2
9	B6	JFK	2
10	UA	EWR	2
11	AA	CLT	2
12	AA	PHX	2
13	WN	SAT	2

In [53]:
```
# TODO: bar chart for bad flights
delay_flight=list(freq['count'])
airline_iata=list(freq.airline_iata_icao)
dest_iata=list(freq.destination_iata_icao)
pos_df=pd.DataFrame({'air':airline_iata,'dest':dest_iata})

pos=[]
for i in range(len(pos_df)):
    a=pos_df.ix[i,'air']
    d=pos_df.ix[i,'dest']
    pos.append((a,d))



plt.bar(np.arange(len(delay_flight)), delay_flight, align='center', alpha=0.5)
plt.xticks(np.arange(len(delay_flight)), pos,rotation='vertical')
plt.ylabel('counts')
plt.title('bar chart')

plt.show()
```



0.12 Step 5.2.4

```
In [ ]: # TODO: delayed flights by airline
        n_df=flights_s_df[(flights_s_df.ArrDelayMinutes >= 30)]
        n_df
```

```
In [38]: fre_df1=df1.groupby(['airline_iata_icao'])['airline_iata_icao'].agg(
             {"count": len}).sort_values(
             "count", ascending=False).reset_index()
         fre_df1
```

Out[38]:

	airline_iata_icao	count
0	WN	16
1	AA	12
2	B6	10
3	UA	10
4	DL	7
5	F9	3
6	AS	2
7	VX	2

8	HA	1
9	NK	1