



# CIS5200 Term Project Tutorial



**Authors:** Gregory David Mamoyac; Tejas Agara Chandrakumar; Zixiao Fu; Nikita Shendkar;  
Ruchita Shinde

**Instructor:** [Jongwook Woo](#)

**Date:** 12/03/2017

## Lab Tutorial

Zixiao Fu ([zfu2@calstatela.edu](mailto:zfu2@calstatela.edu)), Tejas Agara Chandrakumar ([afnu3@calstatela.edu](mailto:afnu3@calstatela.edu)), Ruchita Shinde ([rshinde@calstatela.edu](mailto:rshinde@calstatela.edu)) Nikita Shendkar ([nshendk@calstatela.edu](mailto:nshendk@calstatela.edu)) Gregory Mamoyac ([gmamoya@calstatela.edu](mailto:gmamoya@calstatela.edu))

12/03/2017

# Analysis of Los Angeles Police Department (LAPD) Activity with Demographic and Economic Implications

---

## Objectives

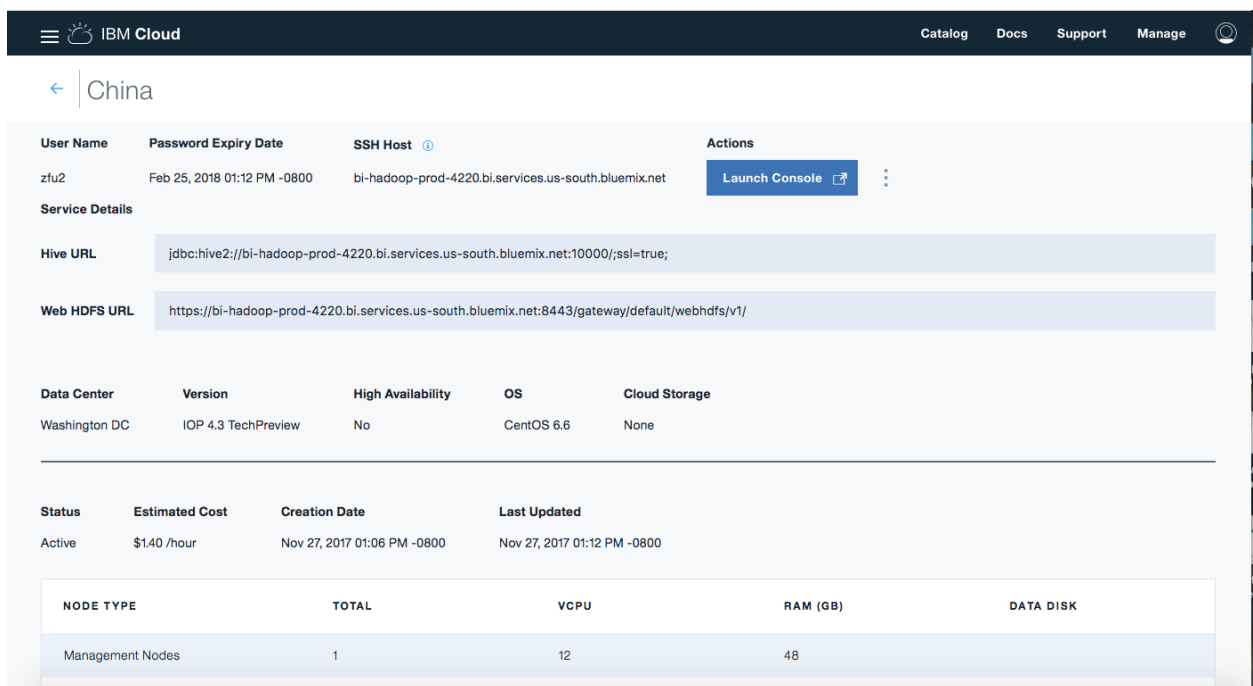
- Get data from government open source at DataLA and ControlPanel LA
- Create Hadoop cluster
- Data ETL written in Pig Latin
- Visualization using Tableau

## Platform Spec

- IBM Bluemix BigInsights
- CPU Speed: Management node 12 VCPU~3GHz, Data node 4 VCPU~3GHz
- # of CPU cores: Management node 12, Data node 4
- # of nodes: Management node 1, Data node 3
- Total Memory Size: Management node 48GB, Data node 1 TB SATA
- Tableau v 10.4

## Step 1: Create cluster using IBM bluemix

1. Sign into your Bluemix account. <https://console.bluemix.net/catalog/>
2. Create a cluster in IBM Bluemix BigInsights
3. Give the cluster name and click create cluster.
4. Chose IOP 4.3 TechPreview version and select Sqoop and Pig function



The screenshot displays the IBM Cloud console interface for a BigInsights cluster. The top navigation bar includes the IBM Cloud logo and links for Catalog, Docs, Support, and Manage. The main content area shows the cluster details for a region set to 'China'.

**Cluster Details:**

- User Name:** zfu2
- Password Expiry Date:** Feb 25, 2018 01:12 PM -0800
- SSH Host:** bi-hadoop-prod-4220.bi.services.us-south.bluemix.net
- Actions:** Launch Console

**Service Details:**

- Hive URL:** jdbc:hive2://bi-hadoop-prod-4220.bi.services.us-south.bluemix.net:10000/?ssl=true;
- Web HDFS URL:** https://bi-hadoop-prod-4220.bi.services.us-south.bluemix.net:8443/gateway/default/webhdfs/v1/

**Configuration Table:**

Data Center	Version	High Availability	OS	Cloud Storage
Washington DC	IOP 4.3 TechPreview	No	CentOS 6.6	None

**Status and Cost Table:**

Status	Estimated Cost	Creation Date	Last Updated
Active	\$1.40 /hour	Nov 27, 2017 01:06 PM -0800	Nov 27, 2017 01:12 PM -0800

**Node Configuration Table:**

NODE TYPE	TOTAL	VCPU	RAM (GB)	DATA DISK
Management Nodes	1	12	48	

5. Open your Terminal or Putty window to log in using SSH function

```

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 97x28
[ZixiaodeMacBook-Air:~ zixiaofu$ ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net ]
The authenticity of host 'bi-hadoop-prod-4220.bi.services.us-south.bluemix.net (169.55.88.220)' c
an't be established.
RSA key fingerprint is SHA256:xy+ZGs/jZ2N5W8ZgvGDs33XLJaxxPFE+Druksry7mfM.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'bi-hadoop-prod-4220.bi.services.us-south.bluemix.net,169.55.88.220' (
RSA) to the list of known hosts.
zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net's password:
IBM's internal systems must only be used for conducting IBM's business or for purposes authorized]
by IBM management
Use is subject to audit at any time by IBM management
-bash-4.1$ █

```

## Step 2: Processing, training and selecting data

---

### Load website log data using 'hdfs' function

--Vehicle and Pedestrian Stops

```

$ wget -O stops.csv https://data.lacity.org/api/views/ci25-wgt7/rows.csv?accessType=DOWNLOAD
$ hdfs dfs -put stops.csv
$ hdfs dfs -ls

```

--Arrests

```

$ wget -O arrests.csv https://data.lacity.org/api/views/yr6-6re4/rows.csv?accessType=DOWNLOAD
$ hdfs dfs -put arrests.csv
$ hdfs dfs -ls

```

--Crimes

```

$ wget -O crimes.csv https://data.lacity.org/api/views/y8tr-7khq/rows.csv?accessType=DOWNLOAD
$ hdfs dfs -put crimes.csv
$ hdfs dfs -ls

```

--Demographic and Economic Statistics

```

$ wget -O demo.csv https://controllerdata.lacity.org/api/views/9z75-
k9sy/rows.csv?accessType=DOWNLOAD
$ hdfs dfs -put demo.csv
$ hdfs dfs -ls

```

--2010 Census

```
$ wget -O census.csv https://data.lacity.org/api/views/nxs9-385f/rows.csv?accessType=DOWNLOAD
```

```
$ hdfs dfs -put census.csv
```

```
$ hdfs dfs -ls
```

--Economy Panel

```
$ wget -O econ.csv https://controllerdata.lacity.org/api/views/4ndk-mmc8/rows.csv?accessType=DOWNLOAD
```

```
$ hdfs dfs -put econ.csv
```

```
$ hdfs dfs -ls
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 113x40
HTTP request sent, awaiting response... 200 OK
[Length: unspecified [text/csv]
Saving to: "census.csv"

[ <=> ] 12,183 --.-K/s in 0s

Last-modified header invalid -- time-stamp ignored.
2017-12-07 00:27:19 (1010 MB/s) - "census.csv" saved [12183]

-bash-4.1$ hdfs dfs -put census.csv
-bash-4.1$ hdfs dfs -ls
Found 5 items
-rw-r--r-- 3 zfu2 hdfs 188269252 2017-12-07 00:23 arrests.csv
-rw-r--r-- 3 zfu2 hdfs 12183 2017-12-07 00:27 census.csv
-rw-r--r-- 3 zfu2 hdfs 375144413 2017-12-07 00:26 crimes.csv
-rw-r--r-- 3 zfu2 hdfs 642 2017-12-07 00:26 demo.csv
-rw-r--r-- 3 zfu2 hdfs 557152405 2017-12-07 00:20 stops.csv
-bash-4.1$ wget -O econ.csv https://controllerdata.lacity.org/api/views/4ndk-mmc8/rows.csv?accessType=DOWNLOAD
--2017-12-07 00:27:50-- https://controllerdata.lacity.org/api/views/4ndk-mmc8/rows.csv?accessType=DOWNLOAD
Resolving controllerdata.lacity.org... 52.206.140.205
Connecting to controllerdata.lacity.org[52.206.140.205]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/csv]
Saving to: "econ.csv"

[ <=> ] 6,913,455 4.23M/s in 1.6s

Last-modified header invalid -- time-stamp ignored.
2017-12-07 00:27:52 (4.23 MB/s) - "econ.csv" saved [6913455]

-bash-4.1$ hdfs dfs -put econ.csv
-bash-4.1$ hdfs dfs -ls
Found 6 items
-rw-r--r-- 3 zfu2 hdfs 188269252 2017-12-07 00:23 arrests.csv
-rw-r--r-- 3 zfu2 hdfs 12183 2017-12-07 00:27 census.csv
-rw-r--r-- 3 zfu2 hdfs 375144413 2017-12-07 00:26 crimes.csv
-rw-r--r-- 3 zfu2 hdfs 642 2017-12-07 00:26 demo.csv
-rw-r--r-- 3 zfu2 hdfs 6913455 2017-12-07 00:28 econ.csv
-rw-r--r-- 3 zfu2 hdfs 557152405 2017-12-07 00:20 stops.csv
-bash-4.1$
```

(This step will take about 14 min)

## Execute PigQL to LOAD and STORE data

(Note: Replace the username “zfu2” when execute the code)

--Execute PigQL

```
$ pig
```

--LOAD and then STORE stops

```
grunt> stops = LOAD '/user/zfu2/stops.csv' USING PigStorage(',') AS
(stop_number:chararray,ref_number:chararray,sex:chararray,descent:chararray,descent_desc:chararray,sto
```

```
p_date:chararray,stop_time:chararray,off1_number:chararray,off1_divnum:chararray,div1_desc:chararray,off2_number:chararray,off2_divnum:chararray,div2_desc:chararray,reporting_dist:chararray,stop_type:chararray,post_stop_activity:chararray);
```

```
grunt> STORE stops INTO 'PROJECT/STOPS';
```

```
--LOAD and then STORE arrests
```

```
grunt> arrests = LOAD '/user/zfu2/arrests.csv' USING PigStorage(',') AS (report_id:chararray,arrests_date:chararray,arrests_time:chararray,area_id:chararray,area_name:chararray,reporting_dist:chararray,age:chararray,sex:chararray,descent_code:chararray,charge_group_code:chararray,charge_group_desc:chararray,arrest_type:chararray,charge:chararray,charge_desc:chararray,address:chararray,cross_street:chararray,location:chararray);
```

```
grunt> STORE arrests INTO 'PROJECT/ARRESTS';
```

```
--LOAD and then STORE crimes
```

```
grunt> crimes = LOAD '/user/zfu2/crimes.csv' USING PigStorage(',') AS (dr_number:chararray,crimes_date_reported:chararray,crimes_date:chararray,crimes_time:chararray,area_id:chararray,area_name:chararray,reporting_dist:chararray,crime_code:chararray,mo_code:chararray,victim_age:chararray,victim_sex:chararray,victim_descent:chararray,premise_code:chararray,premise_desc:chararray,weapon_code:chararray,weapon_desc:chararray,status_code:chararray,status_desc:chararray,crime_code1:chararray,crime_code2:chararray,crime_code3:chararray,crime_code4:chararray,address:chararray,cross_street:chararray,location:chararray);
```

```
grunt> STORE crimes INTO 'PROJECT/CRIMES';
```

```
--LOAD and then STORE demo
```

```
grunt> demo = LOAD '/user/zfu2/demo.csv' USING PigStorage(',') AS (fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_per_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_rate:chararray);
```

```
grunt> STORE demo INTO 'PROJECT/DEMO';
```

```
--LOAD and then STORE census
```

```
grunt> census = LOAD '/user/zfu2/census.csv' USING PigStorage(',') AS (zip_code:chararray,population:chararray,median_age:chararray,total_males:chararray,total_females:chararray,total_households:chararray,avg_household_size:chararray);
```

```
grunt> STORE census INTO 'PROJECT/CENSUS';
```

```
--LOAD and then STORE econ
```

```
grunt> econ = LOAD '/user/zfu2/econ.csv' USING PigStorage(',') AS (category:chararray,indicator:chararray,sub_indicator:chararray,locality:chararray,gender:chararray,council_district:chararray,fiscal_year:chararray,fy_qtr:chararray,calender_year:chararray,cy_qtr:chararray,quarter_start:chararray,as_of_date:chararray,unit_of_measure:chararray,unit_text:chararray,value:chararray,source:chararray);
```

```
grunt> STORE econ INTO 'PROJECT/ECON';
```

```

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 113x37
Time      AvgReduceTime  MedianReductime  Alias  Feature  Outputs
job_1512613154167_0006  1      0      3      3      3      0      0      0      econ
MAP_ONLY      hdfs://bi-hadoop-prod-4220.bi.services.us-south.bluemix.net:8020/user/zfu2/PROJECT/ECON,

Input(s):
Successfully read 35786 records (6913865 bytes) from: "/user/zfu2/econ.csv"

Output(s):
Successfully stored 35786 records (6354092 bytes) in: "hdfs://bi-hadoop-prod-4220.bi.services.us-south.bluemix.net:8020/user/zfu2/PROJECT/ECON"

Counters:
Total records written : 35786
Total bytes written : 6354092
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1512613154167_0006

2017-12-07 02:30:51,851 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-07 02:30:51,853 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-07 02:30:51,879 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-07 02:30:51,881 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-07 02:30:51,901 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-07 02:30:51,903 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-07 02:30:51,930 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>

```

## 2. Analysis of Los Angeles Police Department (LAPD) Activity with Demographic and Economic Implications

(2.1- 2.3 LAPD ACTIVITY)

(2.4- 2.5 DEMOGRAPHIC AND ECONOMIC IMPLICATIONS)

### 2.1 Which LAPD Reporting District Has the Most Reports of LAPD Activity

Select Data for Top Five Reporting Districts Having Reports of Stops, Crimes, Arrests

We have created a customized dataset zipcode.csv containing columns Reporting id, Area name, Zipcode. Since we couldn't find the dataset with all these columns during our search. Upload the zipcode.csv directly to the Ambari page.

/ user / zfu2

+ New directory

Browse...

Clear

Upload

zipcode.csv

zfu2

Search File Names

Name	Size	Last Modified	Owner	Group	Permission	Asc	Name
..							
.staging	-	2017-12-08 16:42	zfu2	hdfs	-rwx-----		
.Trash	-	2017-12-08 14:51	zfu2	hdfs	-rwxr-xr-x		
arrests.csv	179.5 MB	2017-12-06 16:23	zfu2	hdfs	-rw-r--r--		
census.csv	11.9 kB	2017-12-06 16:27	zfu2	hdfs	-rw-r--r--		
crimes.csv	357.8 MB	2017-12-08 02:03	zfu2	hdfs	-rw-r--r--		
demo.csv	0.6 kB	2017-12-06 16:26	zfu2	hdfs	-rw-r--r--		
econ.csv	6.6 MB	2017-12-06 16:28	zfu2	hdfs	-rw-r--r--		
OUTPUT	-	2017-12-08 16:22	zfu2	hdfs	-rwxr-xr-x		

```
grunt>zipcode      =      LOAD      '/user/zfu2/zipcode.csv'      USING      PigStorage(',')      AS
(reporting_dist:chararray,area_name:chararray,zip_code:chararray);
```

## 2.1.1 Which reporting district has the most stops activity

```
grunt> a = JOIN stops by report_dist,zipcode by reporting_dist;
```

```
grunt> b = FOREACH a GENERATE stops::report_dist AS reporting_dist , stops::stop_date AS
crimes_date, zipcode::zip_code AS zipcode;
```

```
grunt> dump b;
```

```
grunt> STORE b INTO 'OUTPUT/STOPSFORMAP';
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 110x29
(2198,03/07/2014,91304)
(2198,04/12/2014,91304)
(2198,11/28/2013,91304)
(2199,04/09/2010,91304)
(2199,04/09/2010,91304)
(2199,03/11/2010,91304)
(2199,03/11/2010,91304)
(2199,03/11/2010,91304)
(2199,03/11/2010,91304)
(2199,02/07/2010,91304)
(2199,02/26/2010,91304)
(2199,02/07/2010,91304)
(2199,04/24/2010,91304)
(2199,02/07/2010,91304)
(2199,06/15/2012,91304)
(2199,02/28/2010,91304)
(2199,04/09/2010,91304)
(2199,02/28/2010,91304)
(2199,03/11/2010,91304)
(2199,03/11/2010,91304)
(2199,01/21/2010,91304)
(2199,03/11/2010,91304)
(2199,01/03/2010,91304)
(2199,04/10/2010,91304)
(2199,04/10/2010,91304)
(2199,03/11/2010,91304)
(2199,03/11/2010,91304)
(2199,02/07/2010,91304)
grunt>

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 110x29
Successfully stored 5748249 records (124103874 bytes) in: "hdfs://bi-hadoop-prod-4220.bi.services.us-south.bluemix.net:8020/user/zfu2/OUTPUT/STOPSFORMAP"

Counters:
Total records written : 5748249
Total bytes written : 124103874
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1512777566506_0034

2017-12-09 19:54:14,091 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-09 19:54:14,093 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-09 19:54:14,123 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-09 19:54:14,128 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-09 19:54:14,167 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-09 19:54:14,170 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-09 19:54:14,190 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
```

## 2.1.2 Which reporting district has the most crimes activity

```
grunt> a = JOIN crimes by reporting_dist,zipcode by reporting_dist;
```

```
grunt> b = FOREACH a GENERATE crimes::reporting_dist AS reporting_dist , crimes::crimes_date AS crimes_date, zipcode::zip_code AS zipcode;
```

```
grunt> dump b;
```



```
grunt>STORE b INTO 'OUTPUT/CRIMESDATAFORMAP';
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 86x25  
(2198,06/01/2012,91304)  
(2198,07/06/2014,91304)  
(2198,03/22/2013,91304)  
(2198,09/11/2012,91304)  
(2198,08/20/2013,91304)  
(2198,03/08/2013,91304)  
(2198,04/01/2011,91304)  
(2198,09/11/2014,91304)  
(2198,09/26/2013,91304)  
(2198,07/24/2014,91304)  
(2198,10/09/2014,91304)  
(2198,03/12/2012,91304)  
(2198,08/23/2014,91304)  
(2198,04/01/2011,91304)  
(2198,03/04/2013,91304)  
(2198,03/15/2012,91304)  
(2198,07/02/2013,91304)  
(2198,03/14/2013,91304)  
(2198,04/19/2013,91304)  
(2198,06/19/2013,91304)  
(2198,12/12/2014,91304)  
(2198,04/28/2012,91304)  
(2198,12/01/2014,91304)  
(2198,09/18/2014,91304)  
grunt>
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net — 86x25  
Job DAG:  
job_1512777566506_0003  
  
2017-12-09 00:22:49,133 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050  
2017-12-09 00:22:49,135 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2017-12-09 00:22:49,174 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050  
2017-12-09 00:22:49,180 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2017-12-09 00:22:49,201 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050  
2017-12-09 00:22:49,203 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server  
2017-12-09 00:22:49,222 [main] INFO org.apache.pig.backend.hadoop.executionengine.map ReduceLayer.MapReduceLauncher - Success!  
grunt>
```

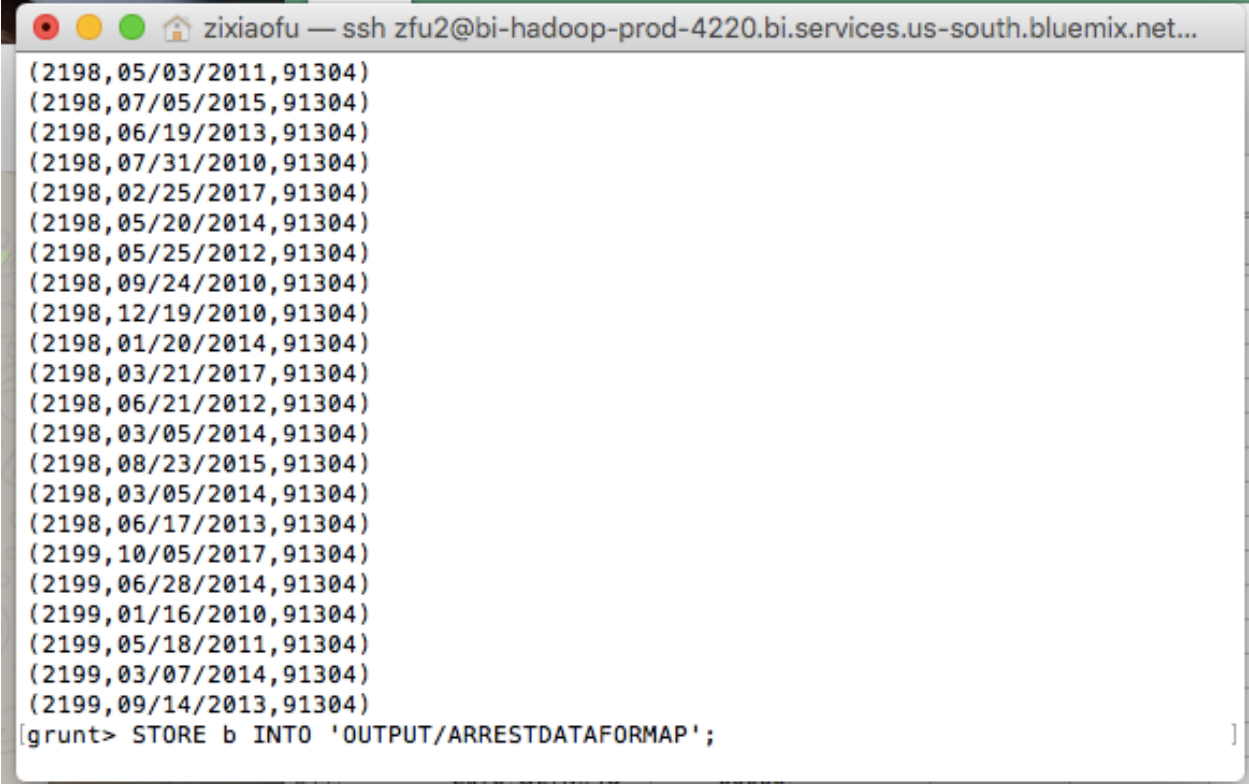
### 2.1.3 Which reporting district has the most arrest activity

```
grunt> a = JOIN arrests by reporting_dist,zipcode by reporting_dist;
```

```
grunt> b = FOREACH a GENERATE arrests::reporting_dist AS reporting_dist , arrests::arrests_date AS  
arrests_date, zipcode::zip_code AS zipcode;
```

```
grunt> dump b;
```

```
grunt> STORE b INTO 'OUTPUT/ARRESTDATAFORMAP';
```



A terminal window titled 'zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...' displays a list of 20 arrest records. Each record is a tuple containing a reporting district, an arrest date, and a zip code. The records are as follows:

- (2198,05/03/2011,91304)
- (2198,07/05/2015,91304)
- (2198,06/19/2013,91304)
- (2198,07/31/2010,91304)
- (2198,02/25/2017,91304)
- (2198,05/20/2014,91304)
- (2198,05/25/2012,91304)
- (2198,09/24/2010,91304)
- (2198,12/19/2010,91304)
- (2198,01/20/2014,91304)
- (2198,03/21/2017,91304)
- (2198,06/21/2012,91304)
- (2198,03/05/2014,91304)
- (2198,08/23/2015,91304)
- (2198,03/05/2014,91304)
- (2198,06/17/2013,91304)
- (2199,10/05/2017,91304)
- (2199,06/28/2014,91304)
- (2199,01/16/2010,91304)
- (2199,05/18/2011,91304)
- (2199,03/07/2014,91304)
- (2199,09/14/2013,91304)

Below the list, the command `grunt> STORE b INTO 'OUTPUT/ARRESTDATAFORMAP';` is entered, followed by a closing bracket `]` on the same line.

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...

2017-12-08 22:59:03,199 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 22:59:03,203 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 22:59:03,248 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 22:59:03,250 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 22:59:03,268 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 22:59:03,276 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 22:59:03,293 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt> █
```

## 2.2 Which Months Are Most and Least Affected by LAPD Activity

### 2.2.1 Which month has the most and least stops activity

```
grunt> stops = LOAD '/user/zfu2/stops.csv' USING PigStorage(',') AS
(stop_number:chararray,ref_number:chararray,sex:chararray,descent:chararray,descent_desc:chararray,stop_date:chararray,stop_time:chararray,off1_number:chararray,off1_divnum:chararray,div1_desc:chararray,off2_number:chararray,off2_divnum:chararray,div2_desc:chararray,report_dist:chararray,stop_type:chararray,post_stop_activity:chararray);
```

```
grunt> a = STREAM stops THROUGH `tail -n +2` AS
(stop_number:chararray,ref_number:chararray,sex:chararray,descent:chararray,descent_desc:chararray,stop_date:chararray,stop_time:chararray,off1_number:chararray,off1_divnum:chararray,div1_desc:chararray,off2_number:chararray,off2_divnum:chararray,div2_desc:chararray,report_dist:chararray,stop_type:chararray,post_stop_activity:chararray);
```

```
grunt> b = FOREACH a GENERATE stop_number AS stop_number, SUBSTRING(stop_date,0,2) AS date;
```

```
grunt> c = GROUP b by date;
```

```
grunt> d = FOREACH c GENERATE group as date,COUNT_STAR(b.stop_number);
```

```
grunt> dump d;
```

```

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
ne.mapReduceLayer.MapReduceLauncher - Encountered Warning UDF_WARNING_2 1 time(s
).
2017-12-08 04:31:56,972 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 04:31:56,994 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 04:31:57,022 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 04:31:57,022 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(01,542885)
(02,491470)
(03,551315)
(04,502983)
(05,511261)
(06,477868)
(07,523033)
(08,538183)
(09,522377)
(10,506904)
(11,481186)
(12,403255)
(,1)
grunt> █

```

### 2.2.2 Which month has the most and least crimes activity

```

grunt> crimes = LOAD '/user/zfu2/crimes.csv' USING PigStorage(',') AS
(dr_number:chararray,crimes_date_reported:chararray,crimes_date:chararray,crimes_time:chararray,area_
id:chararray,area_name:chararray,reporting_disct:chararray,crime_code:chararray,crime_code_desc:chara
rray,mo_code:chararray,victim_age:chararray,victim_sex:chararray,victim_descent:chararray,premise_co
de:chararray,premise_desc:chararray,weapon_code:chararray,weapon_desc:chararray,status_code:chararr
ay,status_desc:chararray,crime_code1:chararray,crime_code2:chararray,crime_code3:chararray,crime_co
de4:chararray,address:chararray,cross_street:chararray,location:chararray);

```

```

grunt> a = STREAM crimes THROUGH `tail -n +2` AS
(dr_number:chararray,crimes_date_reported:chararray,crimes_date:chararray,crimes_time:chararray,area_
id:chararray,area_name:chararray,reporting_disct:chararray,crime_code:chararray,crime_code_desc:chara
rray,mo_code:chararray,victim_age:chararray,victim_sex:chararray,victim_descent:chararray,premise_co
de:chararray,premise_desc:chararray,weapon_code:chararray,weapon_desc:chararray,status_code:chararr
ay,status_desc:chararray,crime_code1:chararray,crime_code2:chararray,crime_code3:chararray,crime_co
de4:chararray,address:chararray,cross_street:chararray,location:chararray);

```

```

grunt> b = FOREACH a GENERATE dr_number AS dr_number, SUBSTRING(crimes_date,0,2) AS date;
grunt> c = GROUP b by date;
grunt> d = FOREACH c GENERATE group as date,COUNT_STAR(b.dr_number);
grunt> dump d;

```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
2017-12-08 04:35:35,423 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 04:35:35,447 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 04:35:35,447 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 04:35:35,464 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 04:35:35,464 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(01,142691)
(02,123558)
(03,138397)
(04,135254)
(05,141214)
(06,138879)
(07,143444)
(08,143700)
(09,137851)
(10,142035)
(11,115088)
(12,119325)
grunt>
```

### 2.2.3 Which month has the most and least arrests activity

```
grunt> arrests = LOAD '/user/zfu2/arrests.csv' USING PigStorage(',') AS
(report_id:chararray,arrests_date:chararray,arrests_time:chararray,area_id:chararray,area_name:chararray,
reporting_dist:chararray,age:chararray,sex:chararray,descent_code:chararray,charge_group_code:chararra
y,charge_group_desc:chararray,arrest_type:chararray,charge:chararray,charge_desc:chararray,address:cha
rarray,cross_street:chararray,location:chararray);
```

```
grunt> a = STREAM arrests THROUGH `tail -n +2` AS
(report_id:chararray,arrests_date:chararray,arrests_time:chararray,area_id:chararray,area_name:chararray,
reporting_dist:chararray,age:chararray,sex:chararray,descent_code:chararray,charge_group_code:chararra
y,charge_group_desc:chararray,arrest_type:chararray,charge:chararray,charge_desc:chararray,address:cha
rarray,cross_street:chararray,location:chararray);
```

```
grunt> b = FOREACH a GENERATE report_id AS report_id, SUBSTRING(arrests_date,0,2) AS date;
grunt> c = GROUP b by date;
grunt> d = FOREACH c GENERATE group as date,COUNT_STAR(b.report_id);
grunt> dump d;
```



```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
2017-12-08 04:41:52,060 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 04:41:52,089 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 04:41:52,092 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 04:41:52,095 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 04:41:52,095 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(01,96089)
(02,88856)
(03,101618)
(04,95665)
(05,100778)
(06,93562)
(07,99998)
(08,99557)
(09,94568)
(10,92721)
(11,78151)
(12,71988)
grunt>
```

## 2.3 Which year has the most and least stops, arrests and crimes activities? & The Relation of Crimes to Arrests

### 2.3.1 Total number of stops data per year

```
grunt> stops = LOAD '/user/zfu2/stops.csv' USING PigStorage(',') AS
(stop_number:chararray,ref_number:chararray,sex:chararray,descent:chararray,descent_desc:chararray,sto
p_date:chararray,stop_time:chararray,off1_number:chararray,off1_divnum:chararray,div1_desc:chararray
,off2_number:chararray,off2_divnum:chararray,div2_desc:chararray,report_dist:chararray,stop_type:char
array,post_stop_activity:chararray);
```

```
grunt> a = STREAM stops THROUGH `tail -n +2` AS
(stop_number:chararray,ref_number:chararray,sex:chararray,descent:chararray,descent_desc:chararray,sto
p_date:chararray,stop_time:chararray,off1_number:chararray,off1_divnum:chararray,div1_desc:chararray
,off2_number:chararray,off2_divnum:chararray,div2_desc:chararray,report_dist:chararray,stop_type:char
array,post_stop_activity:chararray);
```

```
grunt> b = FOREACH a GENERATE stop_number AS stop_number, SUBSTRING(stop_date,6,10) AS
date;
```

```
grunt> c = GROUP b by date;
```

```
grunt> d = FOREACH c GENERATE group as date,COUNT_STAR(b.stop_number);
```

```
grunt> dump d;
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 05:41:50,010 [main] WARN org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Encountered Warning UDF_WARNING_2 1 time(s
).
2017-12-08 05:41:50,010 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 05:41:50,011 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 05:41:50,014 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 05:41:50,014 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(1900,1)
(2010,549940)
(2011,950636)
(2012,929795)
(2013,899444)
(2014,832809)
(2015,620314)
(2016,627677)
(2017,642104)
(,1)
grunt>
```

### 2.3.2 Total number of crimes data per year

```
grunt> crimes = LOAD '/user/zfu2/crimes.csv' USING PigStorage(',') AS
(dr_number:chararray,crimes_date_reported:chararray,crimes_date:chararray,crimes_time:chararray,area_
id:chararray,area_name:chararray,reporting_disct:chararray,crime_code:chararray,mo_code:chararray,vict
im_age:chararray,victim_sex:chararray,victim_descent:chararray,premise_code:chararray,premise_desc:c
hararray,weapon_code:chararray,weapon_desc:chararray,status_code:chararray,status_desc:chararray,cri
me_code1:chararray,crime_code2:chararray,crime_code3:chararray,crime_code4:chararray,address:chara
rray,cross_street:chararray,location:chararray);

grunt> a = STREAM crimes THROUGH `tail -n +2` AS
(dr_number:chararray,crimes_date_reported:chararray,crimes_date:chararray,crimes_time:chararray,area_
id:chararray,area_name:chararray,reporting_disct:chararray,crime_code:chararray,mo_code:chararray,vict
im_age:chararray,victim_sex:chararray,victim_descent:chararray,premise_code:chararray,premise_desc:c
hararray,weapon_code:chararray,weapon_desc:chararray,status_code:chararray,status_desc:chararray,cri
me_code1:chararray,crime_code2:chararray,crime_code3:chararray,crime_code4:chararray,address:chara
rray,cross_street:chararray,location:chararray);

grunt> b = FOREACH a GENERATE dr_number AS dr_number, SUBSTRING(crimes_date,6,10) AS date;
grunt> c = GROUP b by date;
grunt> d = FOREACH c GENERATE group as date,COUNT_STAR(b.dr_number);
grunt> dump d;
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
ing to job history server
2017-12-08 05:44:28,542 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Con
necting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.n
et/173.16.130.1:8050
2017-12-08 05:44:28,544 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 05:44:28,561 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 05:44:28,562 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 05:44:28,568 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 05:44:28,568 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(2010,208600)
(2011,200200)
(2012,200825)
(2013,191814)
(2014,194527)
(2015,214036)
(2016,223309)
(2017,188125)
grunt>
```

### 2.3.3 Total number of arrests data per year

```
grunt> arrests = LOAD '/user/zfu2/arrests.csv' USING PigStorage(',') AS
(report_id:chararray,arrests_date:chararray,arrests_time:chararray,area_id:chararray,area_name:chararray,
reporting_dist:chararray,age:chararray,sex:chararray,descent_code:chararray,charge_group_code:chararra
y,charge_group_desc:chararray,arrest_type:chararray,charge:chararray,charge_decs:chararray,address:cha
rarray,cross_street:chararray,location:chararray);
```

```
grunt> a = STREAM arrests THROUGH `tail -n +2` AS
(report_id:chararray,arrests_date:chararray,arrests_time:chararray,area_id:chararray,area_name:chararray,
reporting_dist:chararray,age:chararray,sex:chararray,descent_code:chararray,charge_group_code:chararra
y,charge_group_desc:chararray,arrest_type:chararray,charge:chararray,charge_decs:chararray,address:cha
rarray,cross_street:chararray,location:chararray);
```

```
grunt> b = FOREACH a GENERATE report_id AS report_id, SUBSTRING(arrests_date,6,10) AS date;
```

```
grunt> c = GROUP b by date;
```

```
grunt> d = FOREACH c GENERATE group as date,COUNT_STAR(b.report_id);
```

```
grunt> dump d;
```



```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
ing to job history server
2017-12-08 05:46:45,058 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Con
necting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.n
et/173.16.130.1:8050
2017-12-08 05:46:45,060 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 05:46:45,084 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 05:46:45,085 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 05:46:45,098 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 05:46:45,098 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(2010,162478)
(2011,157709)
(2012,163473)
(2013,152898)
(2014,139826)
(2015,126780)
(2016,118664)
(2017,91723)
grunt> █
```

## 2.4 Which Demographic Profile is the Most Prevalent in LAPD Activity

### 2.4.1 Which Descent is Most Affected by LAPD Activity

-- Which descent is most affected by stops data

```
grunt> a = GROUP stops by (descent);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (descent), COUNT($1);
```

```
grunt> dump b;
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 06:25:38,283 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Con
necting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.n
et/173.16.130.1:8050
2017-12-08 06:25:38,286 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 06:25:38,306 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 06:25:38,307 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 06:25:38,311 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 06:25:38,311 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(A,216343)
(B,1492826)
(H,2655519)
(I,4178)
(O,341656)
(W,1342203)
(Descent Code,1)
grunt> █
```

**-- Which descent is most affected by crimes data**

grunt> a = GROUP crimes by (victim\_descent);

grunt> b = FOREACH a GENERATE FLATTEN(group) as (victim\_descent), COUNT(\$1);

grunt> dump b;

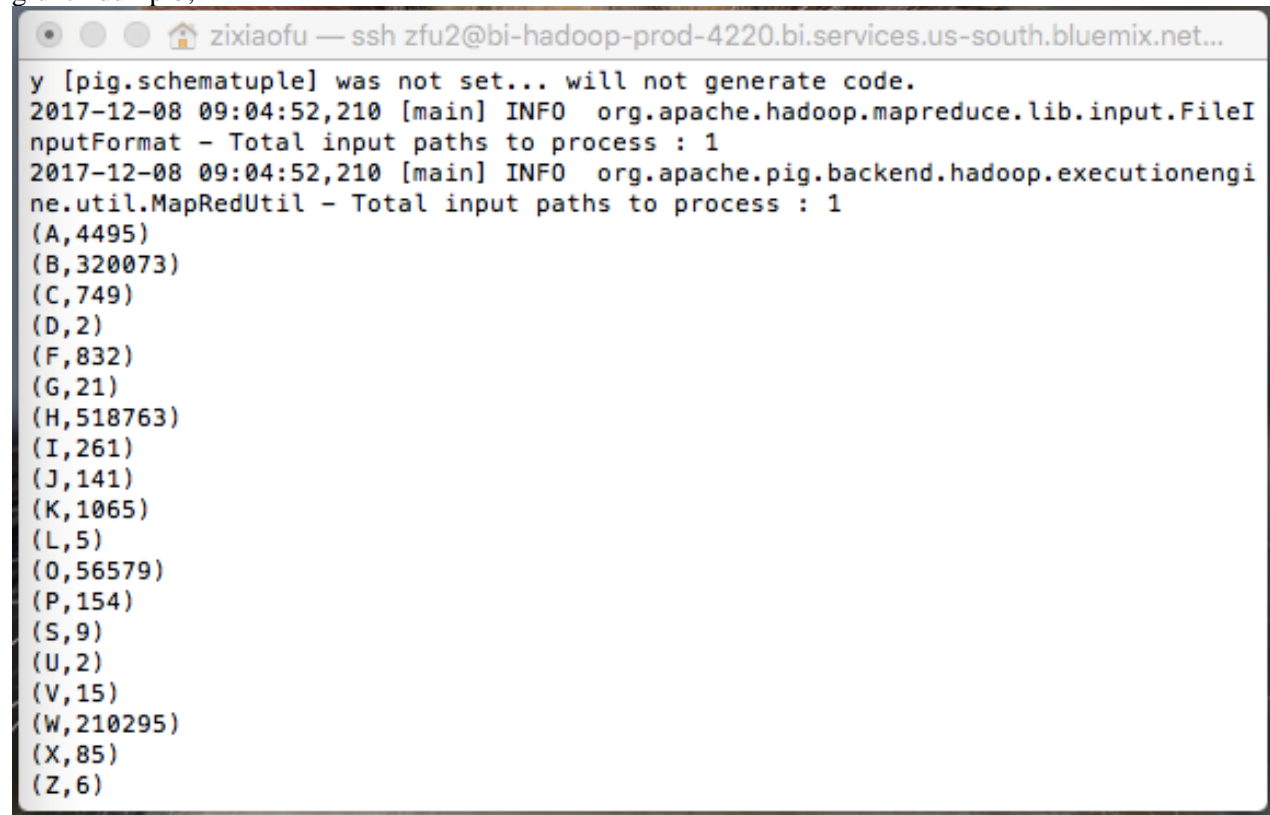
```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
nputFormat - Total input paths to process : 1
2017-12-08 06:36:30,245 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(-,1)
(A,31290)
(B,209475)
(C,511)
(D,13)
(F,94862)
(G,52)
(H,459878)
(I,558)
(J,197)
(K,5961)
(L,8)
(M,123247)
(O,125868)
(P,230)
(S,21)
(U,123)
(V,71)
(W,324020)
(X,38656)
(Z,49)
```

-- Which descent is most affected by arrests data

```
grunt> a = GROUP arrests by (descent_code);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (descent_code), COUNT($1);
```

```
grunt> dump b;
```

A screenshot of a terminal window with a title bar that reads "zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...". The terminal displays the output of a Pig script. It starts with a warning: "y [pig.schematuple] was not set... will not generate code." followed by two INFO messages from Hadoop and Pig about input paths. The main output is a list of pairs (descent\_code, count) for various descent codes: (A, 4495), (B, 320073), (C, 749), (D, 2), (F, 832), (G, 21), (H, 518763), (I, 261), (J, 141), (K, 1065), (L, 5), (O, 56579), (P, 154), (S, 9), (U, 2), (V, 15), (W, 210295), (X, 85), and (Z, 6).

```
y [pig.schematuple] was not set... will not generate code.  
2017-12-08 09:04:52,210 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI  
nputFormat - Total input paths to process : 1  
2017-12-08 09:04:52,210 [main] INFO  org.apache.pig.backend.hadoop.executionengi  
ne.util.MapRedUtil - Total input paths to process : 1  
(A,4495)  
(B,320073)  
(C,749)  
(D,2)  
(F,832)  
(G,21)  
(H,518763)  
(I,261)  
(J,141)  
(K,1065)  
(L,5)  
(O,56579)  
(P,154)  
(S,9)  
(U,2)  
(V,15)  
(W,210295)  
(X,85)  
(Z,6)
```

## 2.4.2 Which Age is Most Affected by LAPD Activity

-- Which age is most affected by crimes data

```
grunt> a = GROUP crimes by (victim_age);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (victim_age), COUNT($1);
```

```
grunt> dump b;
```

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...

```
(97,17)
(96,21)
(95,27)
(94,40)
(93,39)
(92,52)
(91,74)
(90,87)
(89,97)
(88,131)
(87,138)
(86,151)
(85,181)
(84,205)
(83,225)
(82,212)
(81,263)
(80,280)
(79,305)
(78,336)
(77,350)
(76,346)
(75,423)
(74,418)
```

-- Which age is most affected by arrests data

```
grunt> a = GROUP arrests by (age);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (age), COUNT($1);
```

```
grunt> dump b;
```

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...

```
(75,368)
(76,280)
(77,188)
(78,189)
(79,123)
(80,127)
(81,102)
(82,88)
(83,56)
(84,44)
(85,34)
(86,33)
(87,30)
(88,17)
(89,11)
(90,21)
(91,6)
(92,3)
(93,5)
(94,6)
(96,4)
(97,1)
(Age,1)
grunt>
```

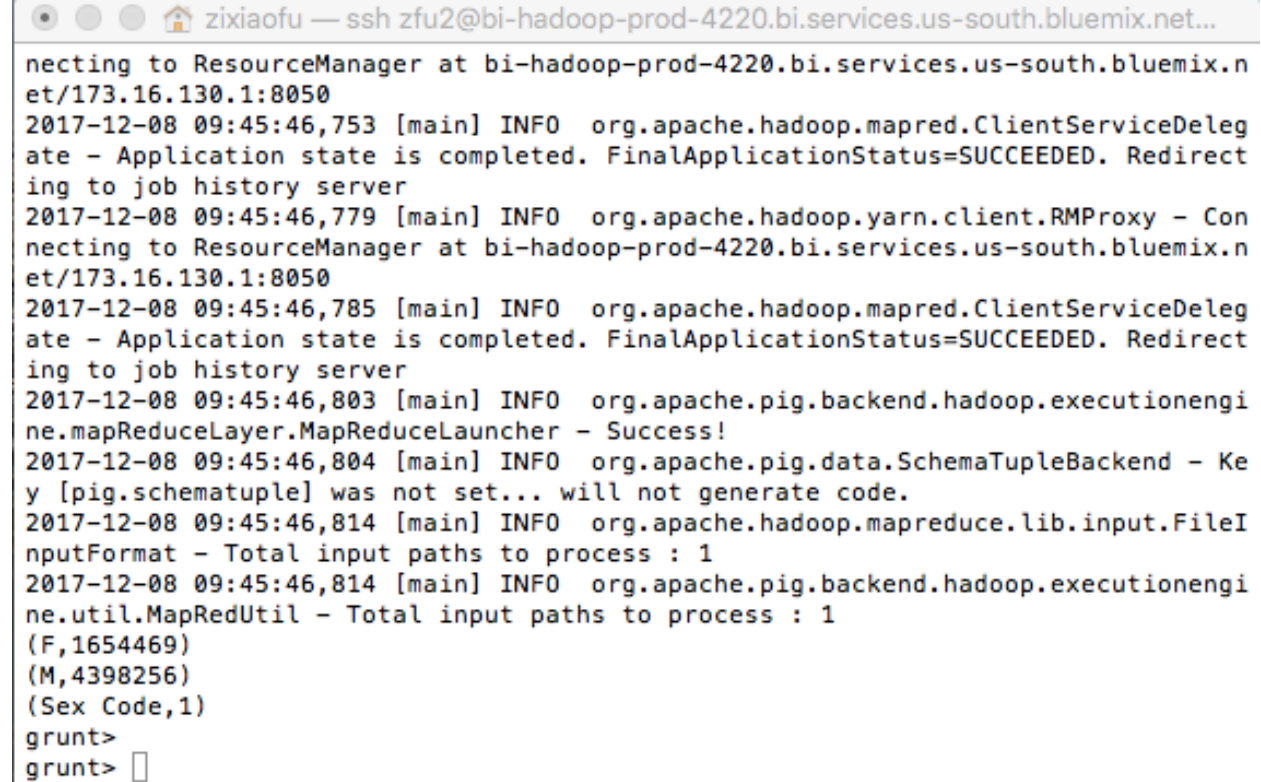
### 2.4.3 Which Gender(sex) is Most Affected by LAPD Activity

-- Which sex is most affected by stops data

```
grunt> a = GROUP stops by (sex);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (sex), COUNT($1);
```

```
grunt> dump b;
```

A terminal window with a title bar showing 'zixiaofu' and an SSH connection to 'ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...'. The terminal displays Hadoop job logs with timestamps and INFO messages from various components like ClientServiceDelegate, RMPProxy, and Pig backend. The logs indicate successful completion of a job. At the bottom, the Grunt commands from the previous block are entered, followed by a cursor.

```
necting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 09:45:46,753 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 09:45:46,779 [main] INFO  org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 09:45:46,785 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 09:45:46,803 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 09:45:46,804 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-12-08 09:45:46,814 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-12-08 09:45:46,814 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(F,1654469)
(M,4398256)
(Sex Code,1)
grunt>
grunt> █
```

-- Which sex is most affected by crimes data

```
grunt> a = GROUP crimes by (victim_sex);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (victim_sex), COUNT($1);
```

```
grunt> dump b;
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
ing to job history server
2017-12-09 19:13:59,064 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Con
necting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.n
et/173.16.130.1:8050
2017-12-09 19:13:59,065 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-09 19:13:59,082 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Con
necting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.n
et/173.16.130.1:8050
2017-12-09 19:13:59,084 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-09 19:13:59,103 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-09 19:13:59,103 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-09 19:13:59,107 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-09 19:13:59,107 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(M,600504)
(F,573820)
grunt> █
```

-- Which sex is most affected by arrests data

```
grunt> a = GROUP arrests by (sex);
```

```
grunt> b = FOREACH a GENERATE FLATTEN(group) as (sex), COUNT($1);
```

```
grunt> dump b;
```



```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
2017-12-08 09:49:31,843 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 09:49:31,846 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 09:49:31,866 [main] INFO org.apache.hadoop.yarn.client.RMPProxy - Connecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluemix.net/173.16.130.1:8050
2017-12-08 09:49:31,867 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2017-12-08 09:49:31,892 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 09:49:31,893 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-12-08 09:49:31,897 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-12-08 09:49:31,897 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(F,231090)
(M,882462)
(Sex Code,1)
grunt> █
```

## 2.5 The Economic Implications of Crimes Per Year

### 2.5.1 The Relation of Personal Income Per Capita to Crimes

```
grunt> demo = LOAD '/user/zfu2/demo.csv' USING PigStorage(',') AS
(fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_per_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_rate:chararray);
```

```
grunt> a = STREAM demo THROUGH `tail -n +2` AS
(fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_per_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_rate:chararray);
```

```
grunt> b = FOREACH a GENERATE fiscal_year AS fiscal_year,income_per_capita AS income_per_capita;
```

```
grunt> c = GROUP b by fiscal_year;
```

```
grunt> d = FOREACH c GENERATE group as fiscal_year,b.income_per_capita;
```

```
grunt> dump d;
```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.net...
et/173.16.130.1:8050
2017-12-08 10:33:22,522 [main] INFO org.apache.hadoop.mapred.ClientServiceDeleg
ate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirect
ing to job history server
2017-12-08 10:33:22,543 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 10:33:22,544 [main] INFO org.apache.pig.data.SchemaTupleBackend - Ke
y [pig.schematuple] was not set... will not generate code.
2017-12-08 10:33:22,547 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2017-12-08 10:33:22,547 [main] INFO org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(2004,{{(34584)}})
(2005,{{(36540)}})
(2014,{{( )}})
( 2006,{{(39508)}})
( 2007,{{(41058)}})
( 2008,{{(42165)}})
( 2009,{{(40396)}})
( 2010,{{(41163)}})
( 2011,{{(43062)}})
( 2012,{{(45800)}})
( 2013,{{(46530)}})
grunt>
```

## 2.5.2 The Relation of Public School Enrollment to Crimes

```
grunt> demo = LOAD '/user/zfu2/demo.csv' USING PigStorage(',') AS
(fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_p
er_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_r
ate:chararray);
```

```
grunt> a = STREAM demo THROUGH `tail -n +2` AS
(fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_p
er_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_r
ate:chararray);
```

```
grunt> b = FOREACH a GENERATE fiscal_year AS fiscal_year,public_school_enrollment AS
public_school_enrollment;
```

```
grunt> c = GROUP b by fiscal_year;
```

```
grunt> d = FOREACH c GENERATE group as fiscal_year,b.public_school_enrollment;
```

```
grunt> dump d;
```



```

zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.ne...
recting to job history server
2017-12-08 10:35:59,089 [main] INFO org.apache.hadoop.yarn.client.RMProxy - C
onnecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluem
ix.net/173.16.130.1:8050
2017-12-08 10:35:59,091 [main] INFO org.apache.hadoop.mapred.ClientServiceDel
egate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redi
recting to job history server
2017-12-08 10:35:59,111 [main] INFO org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 10:35:59,112 [main] INFO org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2017-12-08 10:35:59,117 [main] INFO org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2017-12-08 10:35:59,117 [main] INFO org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(2004, {(900436)})
(2005, {(891252)})
(2014, {(618161)})
( 2006, {(836301)})
( 2007, {(819576)})
( 2008, {(801838)})
( 2009, {(784457)})
( 2010, {(748273)})
( 2011, {(738113)})
( 2012, {(701208)})
( 2013, {(625935)})
grunt> █

```

### 2.5.3 The Relation of Unemployment Rate to Crimes

```

grunt> demo = LOAD '/user/zfu2/demo.csv' USING PigStorage(',') AS
(fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_p
er_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_r
ate:chararray);

```

```

grunt> a = STREAM demo THROUGH `tail -n +2` AS
(fiscal_year:chararray,estimated_population:chararray,income_in_thousands:chararray,income_p
er_capita:chararray,median_age:chararray,public_school_enrollment:chararray,unemployment_r
ate:chararray);

```

```

grunt> b = FOREACH a GENERATE fiscal_year AS fiscal_year,unemployment_rate AS
unemployment_rate;

```

```

grunt> c = GROUP b by fiscal_year;

```

```

grunt> d = FOREACH c GENERATE group as fiscal_year,b.unemployment_rate;

```

```

grunt> dump d;

```

```
zixiaofu — ssh zfu2@bi-hadoop-prod-4220.bi.services.us-south.bluemix.ne...
recting to job history server
2017-12-08 10:42:14,038 [main] INFO org.apache.hadoop.yarn.client.RMProxy - C
onnecting to ResourceManager at bi-hadoop-prod-4220.bi.services.us-south.bluem
ix.net/173.16.130.1:8050
2017-12-08 10:42:14,039 [main] INFO org.apache.hadoop.mapred.ClientServiceDel
egate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redi
recting to job history server
2017-12-08 10:42:14,059 [main] INFO org.apache.pig.backend.hadoop.executionen
gine.mapReduceLayer.MapReduceLauncher - Success!
2017-12-08 10:42:14,060 [main] INFO org.apache.pig.data.SchemaTupleBackend -
Key [pig.schematuple] was not set... will not generate code.
2017-12-08 10:42:14,062 [main] INFO org.apache.hadoop.mapreduce.lib.input.Fil
eInputFormat - Total input paths to process : 1
2017-12-08 10:42:14,062 [main] INFO org.apache.pig.backend.hadoop.executionen
gine.util.MapRedUtil - Total input paths to process : 1
(2004, {(7.2%)})
(2005, {(5.9%)})
(2014, {(9.1%)})
( 2006, {(5.3%)})
( 2007, {(5.6%)})
( 2008, {(8.3%)})
( 2009, {(12.8%)})
( 2010, {(13.9%)})
( 2011, {(13.5%)})
( 2012, {(12.1%)})
( 2013, {(11.3%)})
grunt> █
```

## Step 3: Visualization using Tableau

Download datasets from Ambari, for example.

Ambari China 0 ops 0 alerts Dashboard Services Hosts Alerts zfu2

/ user / zfu2 / PROJECT / ARRESTS

+ New directory Upload

ARRESTS

Search File Names

Name	Size	Last Modified	Owner	Group	Permission	Actions
..						
_SUCCESS	0.1 KB	2017-12-06 18:26	zfu2	hdfs	-rw-r--r--	Download Copy Delete
part-m-00000	118.0 MB	2017-12-06 18:26	zfu2	hdfs	-rw-r--r--	Download Copy Delete
part-m-00001	47.5 MB	2017-12-06 18:26	zfu2	hdfs	-rw-r--r--	Download Copy Delete

Transfer the 'part-m-00000' to txt. file, then open the txt. file in excel and save in excel format.

Open tableau, open the excel file which contain the data after processing. Then do the data visualization works.

Tableau - Book1

Connect

To a File

- Excel
- Text file
- Statistical file
- More...

To a Server

- Tableau Server
- Microsoft SQL Server
- MySQL
- Oracle
- Amazon Redshift
- More...

Saved Data Sources

- Sample - Superstore
- World Indicators

Open

World Indicators

Regional

3131133Using ...

zfu2hw Manipul...

Manipulating Str...

Digital music rev...

676767676Cal...

3

Creating\_A\_Hist...

Sample Workbooks

More Samples

Discover

Training

View all 87 training videos

Viz of the Week

Pedigree Of The Teams Qualified For World Cup 2018

Resources

Blog - Singapore's next generation of data talent

Tableau Conference - Register Now

Forums

Update to 10.4.1 now

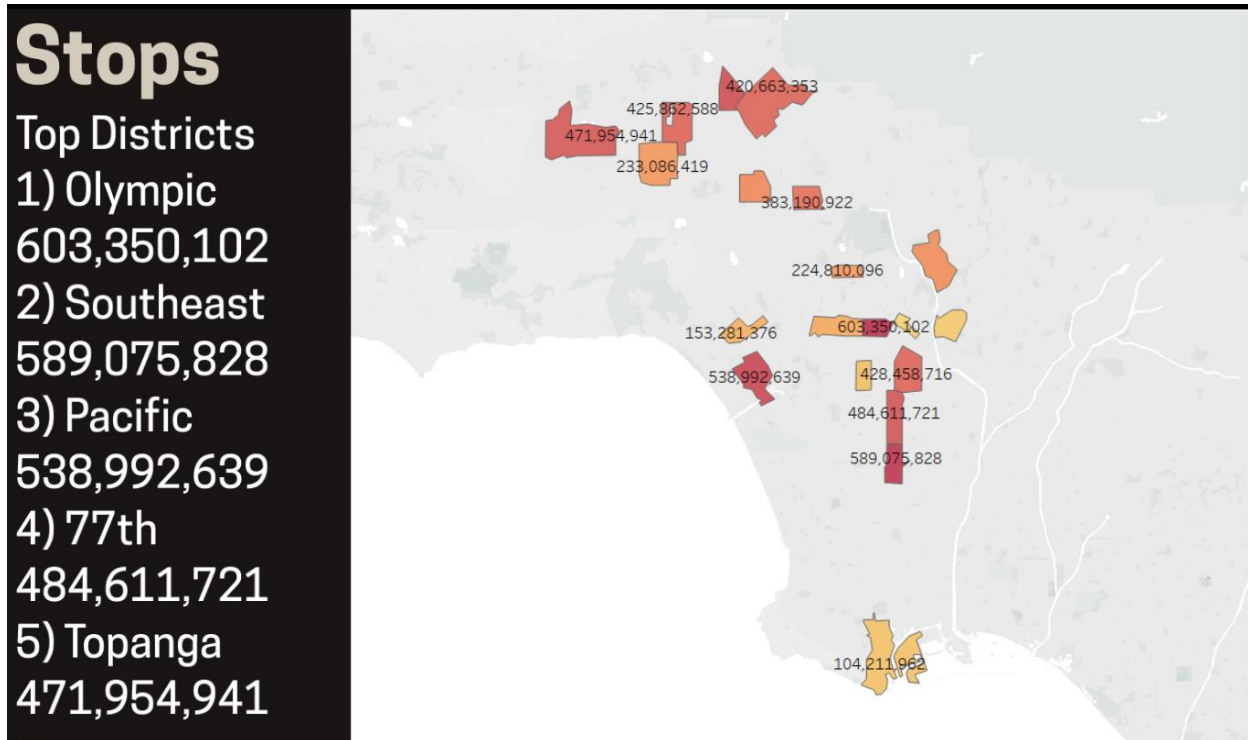
(2.1- 2.3 LAPD ACTIVITY)

(2.4- 2.5 DEMOGRAPHIC AND ECONOMIC IMPLICATIONS)

## 2.1 Which LAPD Reporting District Has the Most Reports of LAPD Activity

To visualize location type of results on map, convert csv file to excel and click on map button under insert tab.

### 2.1.1 Which reporting district has the most stops activity

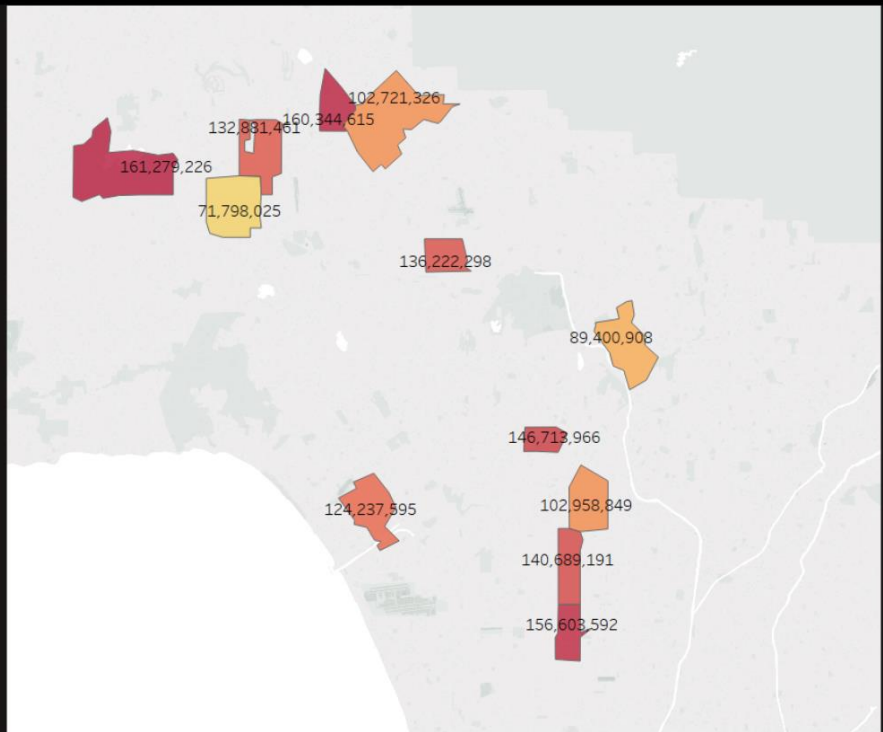


### 2.1.2 Which reporting district has the most crimes activity

# Crimes

## Top Districts

- 1) Topanga  
161,279,226
- 2) Mission  
160,344,615
- 3) Southeast  
156,603,592
- 4) Olympic  
146,713,966
- 5) 77th  
140,689,191

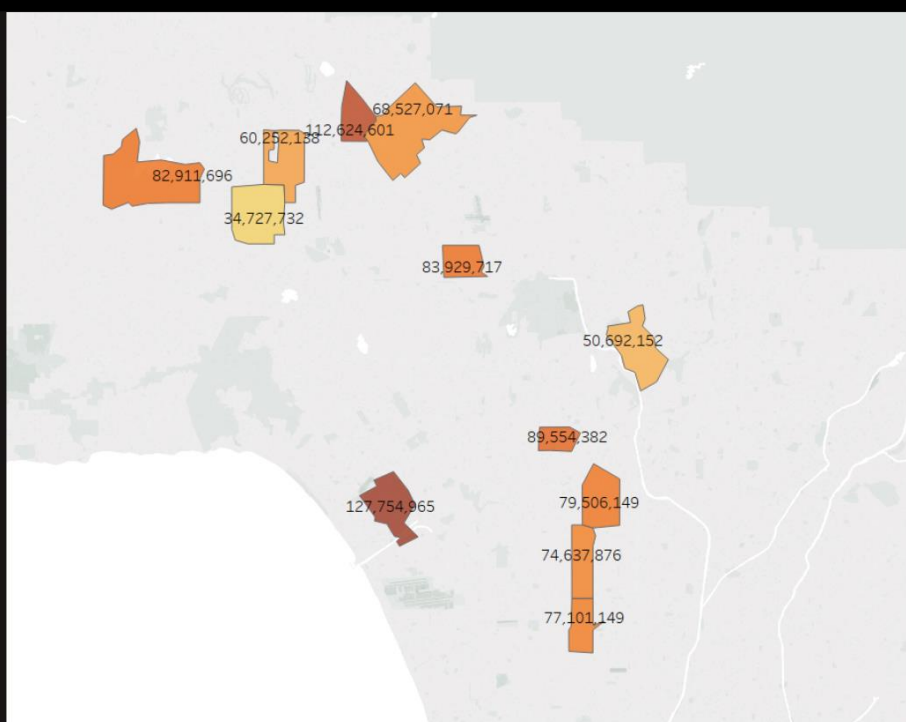


## 2.1.3 Which reporting district has the most arrest activity

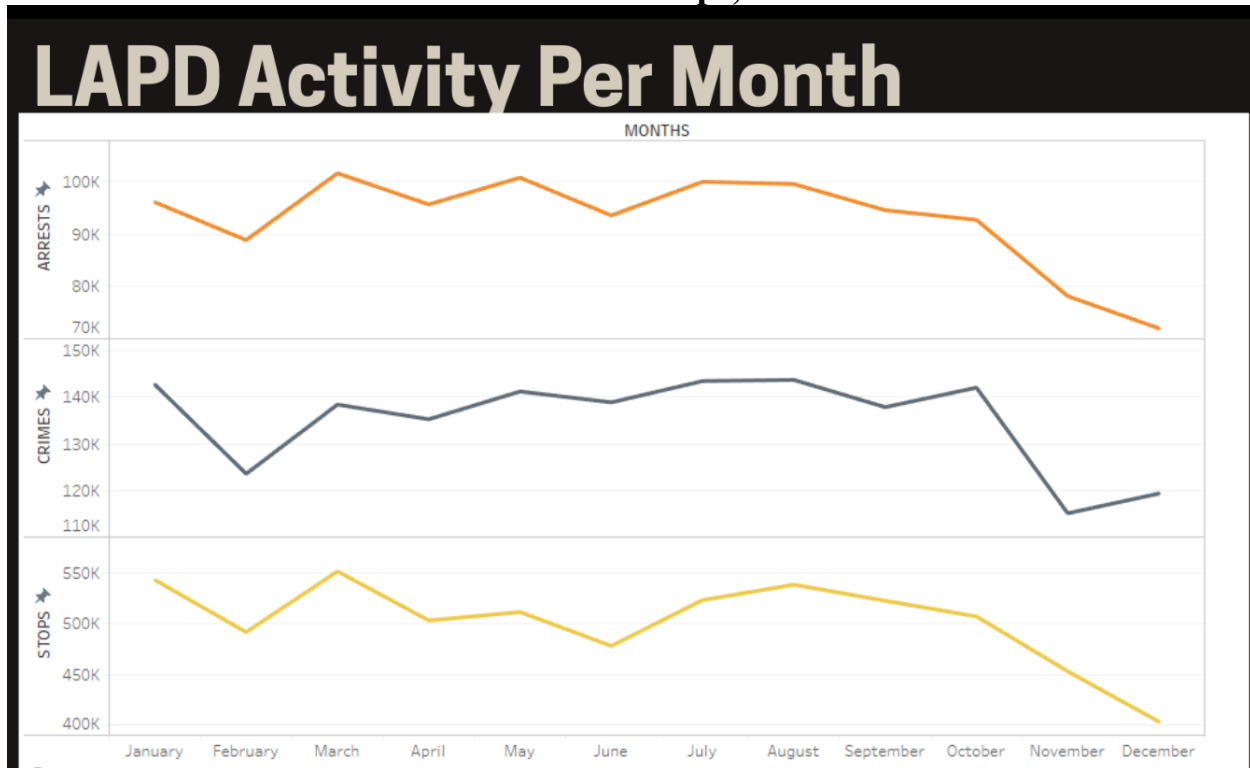
# Arrests

## Top Districts

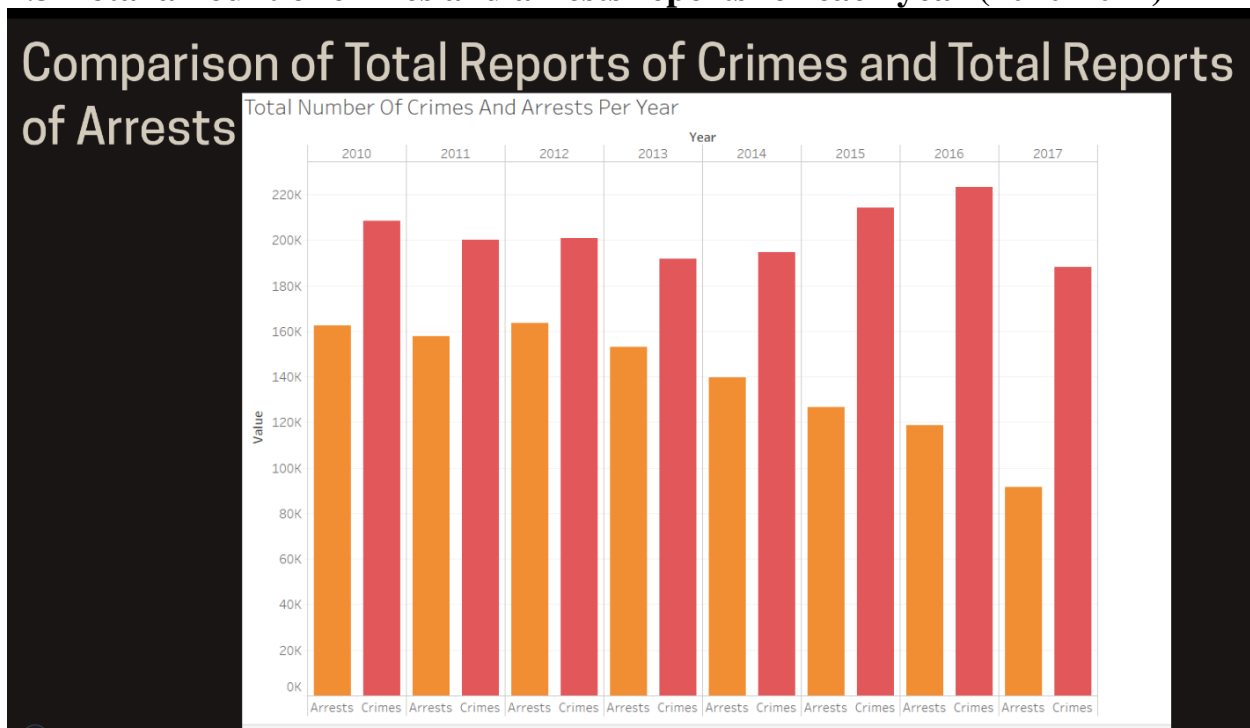
- 1) Pacific  
127,754,965
- 2) Mission  
112,624,601
- 3) Olympic  
89,554,382
- 4) N. Hollywood  
83,929,717
- 5) Topanga  
82,911,696



## 2.2 Which month has the most and least stops, arrests and crimes activities?

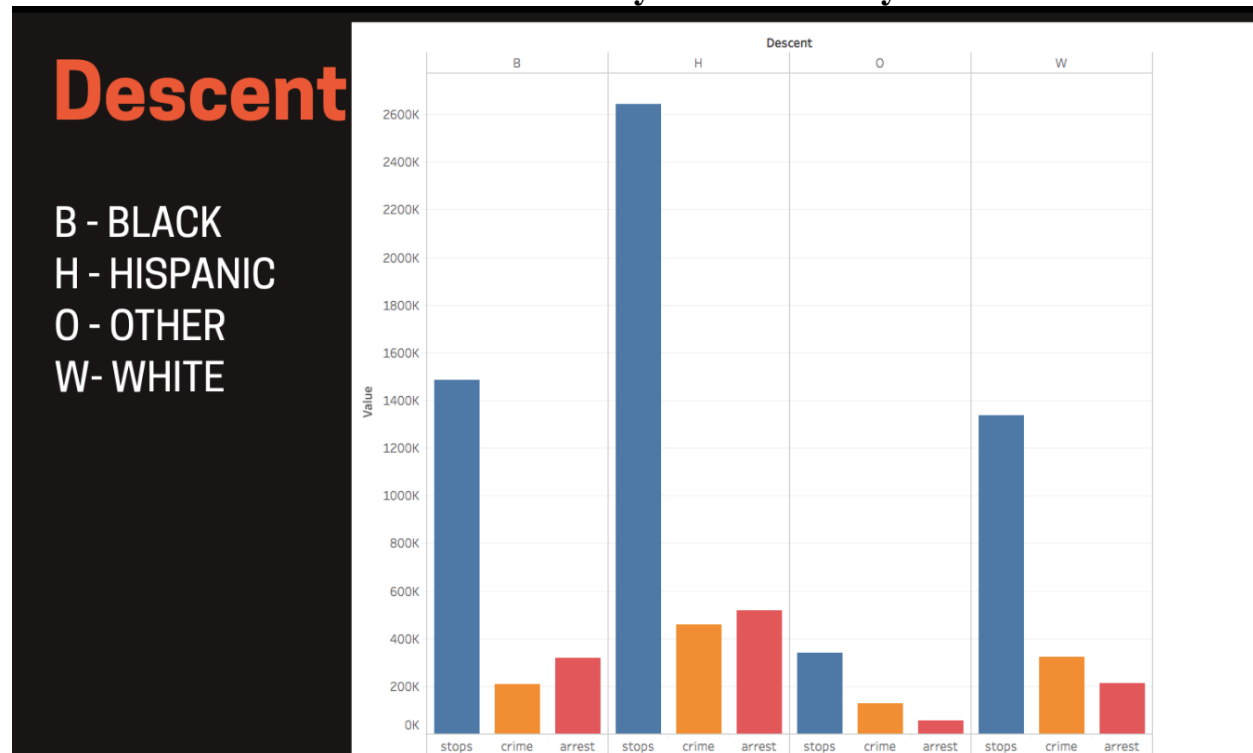


## 2.3 Total amount of crimes and arrests reports for each year (2010-2017)

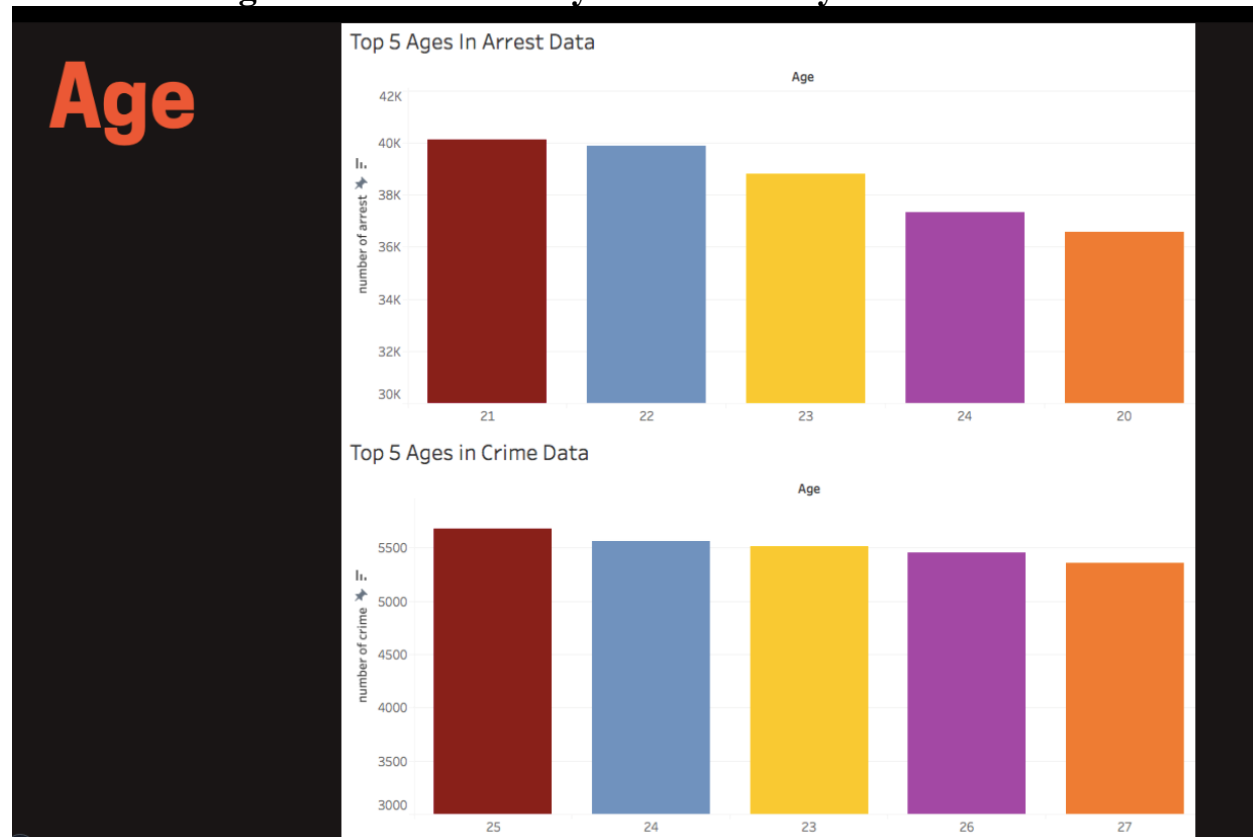


## 2.4 Which Demographic Profile is the Most Prevalent in LAPD Activity

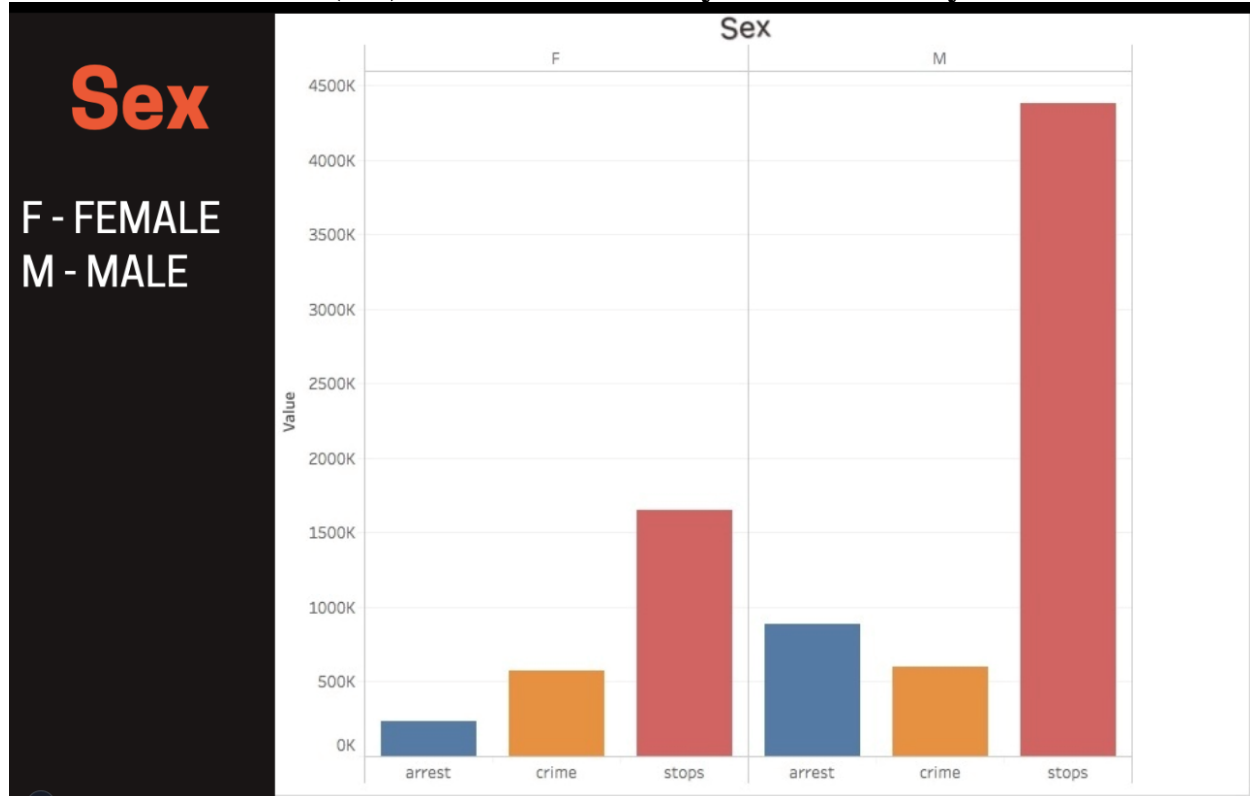
### 2.4.1 Which Descent is Most Affected by LAPD Activity



### 2.4.2 Which Age is Most Affected by LAPD Activity



### 2.4.3 Which Gender(sex) is Most Affected by LAPD Activity

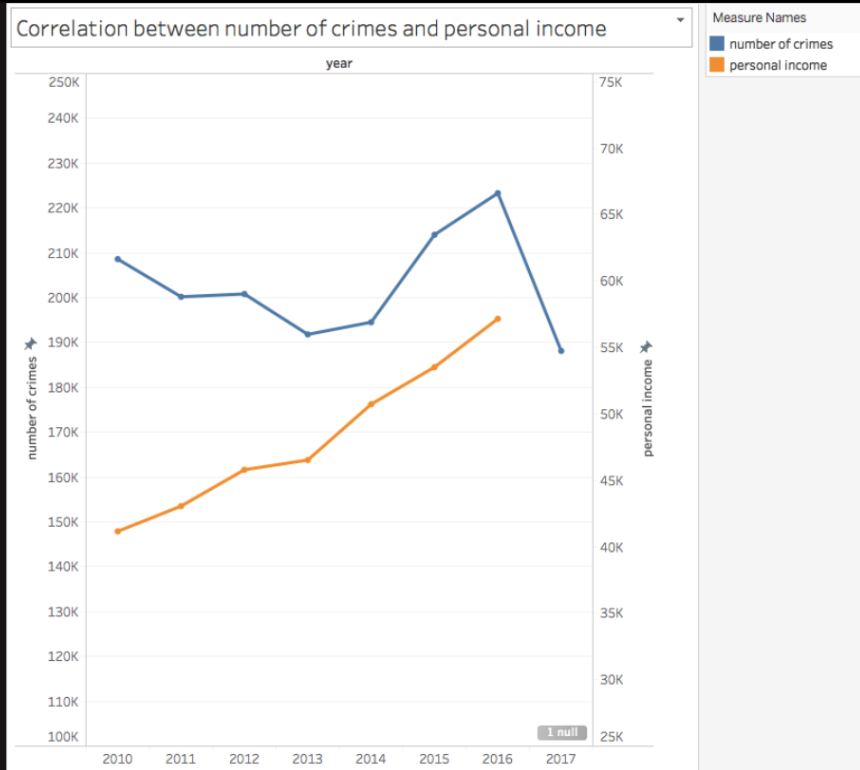


### 2.5 The Economic Implications of Crimes Per Year

#### 2.5.1 The Relation of Personal Income Per Capita to Crimes

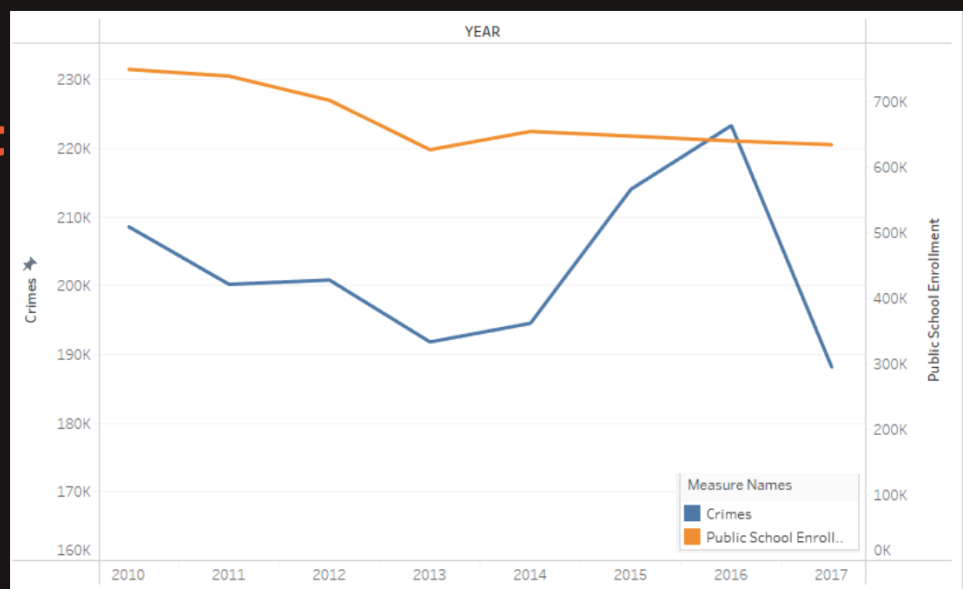


## The Relation of Personal Income Per Capita to Crimes



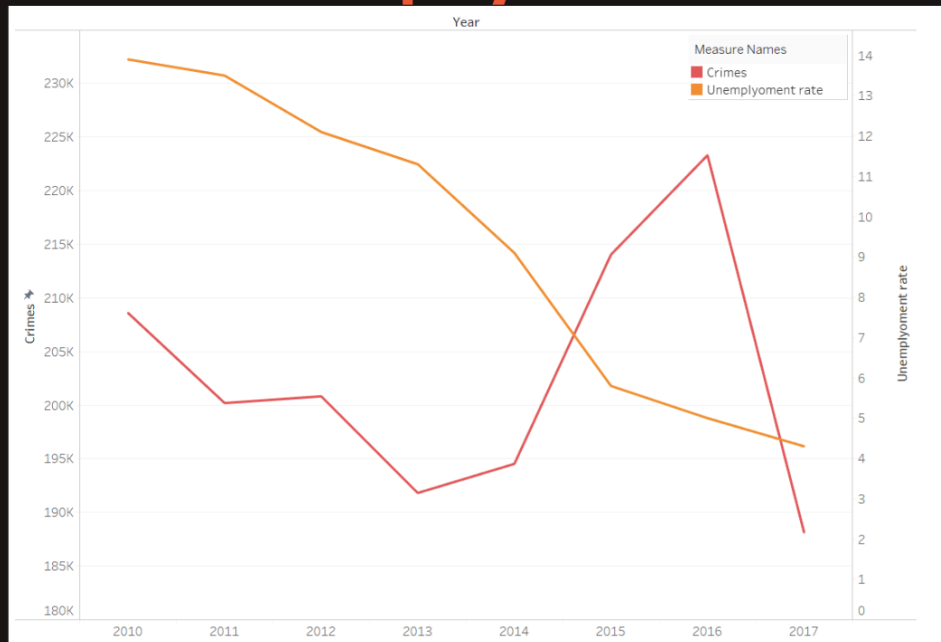
### 2.5.2 The Relation of Public School Enrollment to Crimes

## The Relation of Public School Enrollment to Crimes



### 2.5.3 The Relation of Unemployment Rate to Crimes

## The Relation of Unemployment Rate to Crimes



## References

1. URL of Data Source, <https://data.lacity.org/browse?tags=lapd>
2. URL of your Github, <https://github.com/acteias/Project-CIS5200>