Checkpoint1: SQL Analytics

Team members:
Hong Hong (NetID: hha7337)
Hui Ye (NetID:hyr1096)

Our group attempts to conduct research on the diversity of police officers and discuss the relationship between policing diversity and misconduct rate, including race and gender representativeness. In this checkpoint, our group analyzed the following 5 questions using SQL, the corresponding SQL scripts and guide on how to run the script are attached to the folder.

1. **What's the average police misconduct rate by area?**
   The main purpose of raising this question is to pave the way for subsequent analysis and data exploration. Since the misconduct rate may be continuous data, in this question, we hope to find an average, or a distinguishing point, in the following In the problem, we were able to use this average to rank the severity of the miscount rate into categorical data. We could use this average to divide the severity of the misconduct into server misconduct, medium misconduct and mild misconduct.

   To examine this question on SQL, we first decide to use *police-districts* as our area to quantify the question instead of the *community* or *neighborhood.* The thing that we could conduct the sql successfully is we need to explore the relation between the exact police officer with a certain area. And the way that use *police-districts* to divide the Chicago map is the only way to seek the way out on this relationship.

   The misconduct rate of a district for our question is defined as:
   (Allegation_count / population / police_population) * 1000

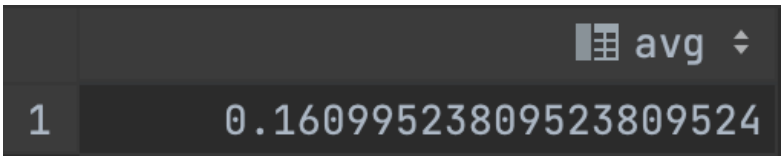

Figure 1

   As shown in the above figure, the average of the misconduct rate by police-districts is around 16%. This finding is actually a bit surprising. It is equivalent to about 16 out of every 100 people in the police patrol area being misconducted by the police. Of course, we cannot rule out factors such as the residents' accusations are not directly equivalent to the police's misconduct, or some residents repeatedly made different misconduct accusations and so on. With this baseline, we can therefore better analyze the following problems. Therefore, we are going to define mild misconduct as 0%~10%, medium misconduct as 11%~25%, and a misconduct rate more than 25% would be considered as severe misconduct.

2. **What's the average proportion of non-white police officers serving in the same area?**
   Since we want to continue exploring racial diversity in a follow-up question, but the concept is a bit vague, in order to make diversity quantifiable and expressible, we decided to take a

simpler approach: we will take the numerical value obtained in this question as the baseline and use those to divide the level of diversity in the composition of police officers.

avg_nonwhite	avg_white	avg_black	avg_hispanic	avg_asian
0.4121791666666666667	0.5878208333333333333	0.2386916666666666667	0.1527833333333333333	0.0184625

Figure 2

For the race aspect, we can see from figure 2 that in the police force of each same patrol range, the average proportion of all white police officers is as high as 58.8%, while the average proportion of the remaining non-white police officers is only 41.2%. This ratio is actually quite amazing. We can understand that the average police force in an area is more than half of the white police.

At the same time, among the non-white police, black police have the highest proportion, reaching 23.9%, followed by hispanic accounting for 15.3%, and the proportion of Asian police is the lowest, only 2%.

avg_female	avg_male
0.17894166666666666667	0.82105833333333333333

Figure 3

To explore a deeper level view of this problem, we conduct the query on the gender aspect as well. We could tell from figure 3 that the average female composition in policing is extremely small compared to that of male, 17.9% versus 82.1% respectively.

Therefore, we define low diversity when the proportion of non-white police officers is 0%～30%. When the proportion of non-white police officers is 31%～55%, it is medium diversity, and when the proportion of non-white police officers is higher than 55%, it is high diversity.

3. **Which two areas have similar race demography? Find 5 typical groups.**

To simplify this problem, we will use control variables to make the division of similar demography. We will divide the similar demography by filtering out communities with different ethnic groups as the main components, and then comparing the percentages of the remaining ethnic groups.

Black-dominant district:

district	total	asian	black ▼ 1	white	hispanic	other
11	90841	0.0007	0.9745	0.0034	0.0101	0.0095
8	71071	0.0008	0.9679	0.0037	0.0164	0.0093
9	75235	0.0041	0.9438	0.0196	0.0149	0.0153
14	74396	0.0005	0.9418	0.0113	0.0339	0.0109
21	59458	0.0029	0.9329	0.0161	0.0381	0.0082
20	70474	0.0047	0.8467	0.0236	0.1152	0.0088
25	95439	0.0612	0.6915	0.186	0.034	0.026
10	123575	0.0021	0.6182	0.0803	0.2863	0.0112

Figure 4

We sort the proportion of black residents in descending order, and in the first few districts with higher proportion, we can find that district 8 and 11, district 9 and 14 have similar race demography. Here, we comprehensively consider the proportions of other races. When the

absolute difference between the proportions of other races does not exceed 10%, we consider them to constitute similar demography.

White-dominant district:

district	total	asian	black	white ▼ 1	hispanic	other
18	117041	0.0835	0.0913	0.7554	0.0496	0.0192
5	200786	0.0611	0.0663	0.7498	0.0997	0.0215
16	199482	0.0537	0.0101	0.692	0.2278	0.0149
2	91279	0.1401	0.1086	0.5414	0.181	0.0265
23	62781	0.1719	0.2143	0.5249	0.06	0.0275
15	141038	0.1474	0.1795	0.4289	0.2109	0.0307
7	117738	0.0306	0.0682	0.423	0.459	0.0176
19	127869	0.0637	0.1802	0.4049	0.331	0.0187
1	144096	0.1206	0.0332	0.3868	0.4319	0.0253

Figure 5

From figure 5, we can see that district 5 and 18 have similar race demography when we consider them as white-dominant districts.

Hispanic-dominant district:

district	total	asian	black	white	hispanic ▼ 1	other
6	200391	0.0154	0.1648	0.1466	0.6637	0.0084
22	118093	0.002	0.334	0.0364	0.6219	0.0045
24	165201	0.1567	0.1153	0.1471	0.5727	0.0071
17	247373	0.0081	0.2111	0.2082	0.5654	0.0063
7	117738	0.0306	0.0682	0.423	0.459	0.0176
1	144096	0.1206	0.0332	0.3868	0.4319	0.0253
19	127869	0.0637	0.1802	0.4049	0.331	0.0187
10	123575	0.0021	0.6182	0.0803	0.2863	0.0112

Figure 6

From figure 6, we can see that district 6 and 24 have similar race demography when we consider them as hispanic-dominant districts.

Asian-dominant district:

district	total	asian ▼ 1	black	white	hispanic	other
23	62781	0.1719	0.2143	0.5249	0.06	0.0275
24	165201	0.1567	0.1153	0.1471	0.5727	0.0071
15	141038	0.1474	0.1795	0.4289	0.2109	0.0307
2	91279	0.1401	0.1086	0.5414	0.181	0.0265
1	144096	0.1206	0.0332	0.3868	0.4319	0.0253
18	117041	0.0835	0.0913	0.7554	0.0496	0.0192
19	127869	0.0637	0.1802	0.4049	0.331	0.0187

Figure 7

In fact, from here, we can find that there are no districts mainly populated by Asians. So we just repeat the previous sorting step and pick out that district 2 and 15 have similar race demography.

Therefore, in this problem we obtain five groups of districts with similar race demography. They are district 8 and 11, district 9 and 14, district 5 and 18, district 6 and 24, district 2 and 15. The communities dominated by different residential groups also show different characteristics.

In the black-dominant district, the proportion of black residents is as high as 97%, which means that the total proportion of the remaining groups will not exceed 3%. Among them, hispanic can account for the remaining 1%, and the remaining white and asian residents account for a very small proportion. This data shows that the racial composition of black-majority communities is relatively uniform.

In the white-dominant districts, the highest proportion of white residents is only 75%, which means that there are 25% of other racial populations remaining in the same district. Similarly, in hispanic-dominant districts, the highest percentage of Latino residents is only 66%. Similar to white-dominant districts, these communities have richer race demography.

Considering that the Asian population in the Chicago area may be relatively small, we cannot find a district where Asians live mainly, because the highest proportion is only 17%, which is still very different from other races. But one of the interesting findings is that we can see that areas with a relatively high proportion of Asian residents also have a higher proportion of white residents. In figure 7, the average proportion of white residents is more than 50%.

4. **What's the proportion of non-white police officers who are policing the above areas?**
   To delve deeper into our topic, we will explore the proportion of non-white in the police force in policing-related areas in this question.

	district	non_white	black	white	hispanic	female	district_misconduct_rate
1	8	0.2346	0.0603	0.7654	0.162	0.1441	0.1726
2	11	0.4249	0.2409	0.5751	0.1678	0.1759	0.1509

Figure 8 Black-dominant district 8 and 11

	district	non_white	black	white	hispanic	female	district_misconduct_rate
1	9	0.2952	0.0855	0.7048	0.189	0.1752	0.1596
2	14	0.4439	0.0697	0.5561	0.3367	0.1684	0.1905

Figure 9 Black-dominant district 9 and 14

	district	non_white	black	white	hispanic	female	district_misconduct_rate
1	2	0.7816	0.7038	0.2184	0.0627	0.223	0.046
2	15	0.4587	0.2698	0.5413	0.1603	0.1841	0.06

Figure 10 District 2 and 15

	district	non_white	black	white	hispanic	female	district_misconduct_rate
1	18	0.3156	0.197	0.6844	0.0947	0.124	0.0756
2	5	0.6243	0.5573	0.3757	0.0587	0.2626	0.0609

Figure 11 White-dominant district 5 and 18

	district	non_white	black	white	hispanic	female	district_misconduct_rate
1	24	0.2491	0.0596	0.7509	0.1318	0.1823	0.1014
2	6	0.6936	0.5554	0.3064	0.1258	0.2229	0.0509

Figure 12 Hispanic-dominant district 6 and 24

The exact proportions are shown in the above figure(8~12). To gain more insights, we can see that in communities where black residents are the majority, the police officers who are policing the most are white police officers, with an average of more than 65%, while non-whites can only account for a minority. In districts 9 and 14 with more complicated race composition, the proportion of non-white police officers is relatively higher.

Taking gender into account, we can see some slight differences from the graph. The proportion of female police officers is relatively low in the black-dominant area. In white-dominant districts, the proportion of female police officers can reach 26%. This is actually a quite surprising proportion, because we can find from question 2 that the average proportion of female police officers is only 17%.

5. **What's the misconduct rate in these areas?**

The exact misconduct rates are also shown in the above figure(8~12). From which we could see the huge difference between the misconduct rates on different groups of districts. Among them, the average misconduct rate of black-dominant districts is around 17%, and the highest can go to 19%. In our division, this misconduct rate can be called medium severity of misconduct. On the other hand, for the remaining groups, the misconduct rate generally does not exceed 10%, which belongs to the mild severity of misconduct.

At the same time, there is also an interesting finding. When we compare the five groups, we can find that when the proportion of female police officers in one district of the same type is higher than that of the other district, the misconduct rate will be relatively lower. When we tried to explore the influence of the proportion of non-white police on misconduct rate, we could not find the law of consistency in all five groups. The relationship shown in figure 9, the higher the proportion of non-white police, the higher the proportion of non-white police.