Glorious Gorillas
Findings

Data Cleaning and Integration:

The theme of the project is investigating the trends between the seniority of an investigator in terms of their age and rank, and their history of complaints. The focus of this investigation is twofold: First, how does the history of complaints correlate with each investigator over time? For example, as an investigator rises in the ranks, does the rate of complaints increase for this officer? Similarly, is there a trend in the type of complaint that the investigator receives? Second, we can understand potential biases that the officers might hold against their victims.

Additionally, investigating this topic may uncover related trends that may not necessarily be confined to the scope of this topic, but are interesting and important nonetheless. For example, trends between officers that rise in rank and their complaint volume might reveal some information about how the Chicago PD operates as an institution. That is, an officer may receive complaints from the public after an incident but may be awarded or promoted internally but the institution.

For this checkpoint we wanted to answer 4 questions:

1. How does years worked as a police officer correlate to the amount of arrests they have over their career?
2. Is there any significance between how many complaints an officer has with the amount of arrests they have?
3. Can the data tell us if there are any biases in officer judgments for arrest by understanding arrests as a proportion for each ethnic/racial demographic?
4. Understand how arrests relate to misconduct categories for officers and how many arrests were made for each type of misconduct.

The data cleaning segment was one of the most interesting and intuitive so far, but also rather time consuming. After successfully finishing the checkpoint 3, we can confirm that the cleaning part is the most time consuming part of any data analysis.

For question 1, our group sought to investigate how years worked as a police officer correlate to the amount of arrests they have over their career. Interestingly, there does seem to be a trend ‘bad apples’. There are a select few officers who account for a lot of arrests who have on average over 15 years of work in the field. This contributes to our research question because we can see that officers working longer not only obviously have more arrests as time goes on, but the ‘bad apples’ seem to all be in the late stages of their careers and are more senior in their experience and at times, more senior in their positions. For future directions of inquiry, it would be interesting to understand the rankings and awards for these ‘bad apple’ cops to see how these people progress in their careers with the complaints and arrests they have.

Here is a sample of our results:

	officer_id	years_worked	count_arrests
1	19946	21	3
2	24601	14	136
3	13554	22	58
4	1037	25	1732
5	716	2	0
6	25282	0	0
7	20468	26	0
8	31813	17	483
9	5262	35	0
10	12370	33	0
11	2451	35	0
12	19471	32	0
13	14487	12	88
14	10677	25	0
15	22123	33	0

For question 2, we wanted to understand if there is any significance between how many complaints an officer has with the amount of arrests they have? Interestingly, the amount of arrests made per officer seem to be held closely in tandem with complaints made against officers. In some cases, there are as many arrests as there are complaints. This is a natural follow up to question 1 as this data tell us how many complaints and arrests each officer had. We wanted to expand our research by not only understanding a timeline of years served and arrests but also what kind of story the data can tell around an officer's proportion of arrests with complaints. According to our results, it seems as though there is very much a tight connection. While again there would of course need to be more data and testing to support such a hypothesis, it can be theorized that most people complain about officers when there is an arrest involved. Future lines of inquiry should most certainly seek to understand who these complainants are and if there seems to be any trends amongst people who are seeking out to make complaints.

Here is a sample of our results:

	officer_id	count_arrests	count_compaints
1	1	2754	2766
2	2	7088	7128
3	4	210	238
4	6	582	582
5	7	924	1155
6	8	168	168
7	11	35	49
8	13	380	380
9	14	228	228
10	15	170	209
11	16	6631	7011
12	17	525	525
13	18	1560	1560
14	19	945	960
15	20	12	12

For question 3, we looked at the data to understand any biases in officer judgments for arrest by understanding arrests as a proportion for each ethnic/racial demographic. As shown below, there are 5 racial groups which have counts of complaints from each group as well as the count of arrests. It should be noted that a lot a large portion of this data is a result of not reporting the race of the complainant. When entering this data, this lack of information contributed greatly to having 'dirty data'. However, from the vast amount of data that we as researchers can understand we can see that the number of complaints and arrests strongly resemble one another. A hypothesis can be drawn that a large number of complaints coming in were a result of an arrest due to police misconduct which is a significant finding if proven to be true with more supporting data. This data helped advance our understanding for our research question as we wanted to understand any 'hidden' trends. This certainly seem to show biases for certain demographic groups in the Chicago area. The city of Chicago is made up of 45.3% people of the white demographic and 32% of people in the black demographic group (according to the http://worldpopulationreview.com/us-cities/chicago-population/ census). There seems to be a disproportionate amount of arrests being made for people of color. This unfortunately (and fortunately) sheds light on trends that a minority of Chicago citizens seem to be facing disproportionately. It would certainly be interesting for future directions to understand the demographic groups of people making complaints to see if there is a proportionate amount of people making complaints for wrongful arrests.

Here is a sample of the results: (Note the blank entry in cell 1 denotes unregistered/unknown races)

	race	count_complaints	count_arrests
1		1030465	989915
2	Asian/Pacific Islander	98301	93420
3	Black	5714215	5495519
4	Hispanic	955356	916883
5	Native American/Alaskan Native	8569	8364
6	White	2194154	2106466

For question 4, we wanted to understand how arrests relate to misconduct categories for officers and how many arrests were made for each type of misconduct. We encountered one vital issue here, which was linking between data_arrests, and the copa table where the results below come from. This is an issue still in the works, but we plan on hopefully finding a way to link the sets so that we can also get an arrest count for each complaint. This will give us insight into how many arrests are made for each respective category of crime, which will then in turn allow us to see and formulate potential trends between the category of complaint, and the volume of arrests for that claim. This would also open up to the idea of ranking severity based on categories. Further just to note, there are missing entries within the copa data set which may require more data cleaning down the road. As far as current takeaways from the existing results, we see that the highest amount of complaints go towards excessive force, and having knowledge of the amount of arrests associated with this would be interesting in answering and exploring the questions addressed above.

Here is a sample of the results:

	current_category	category_count
1	Civil Suits	2056
2	Miscellaneous	4143
3	Taser Notification	5401
4	Firearm Discharge at Animal	880
5	*<null>*	0
6	Operational Violation	176
7	Bias	23
8	Motor Vehicle Related Death	68
9	Firearm Discharge – No Hits	517
10	Legal Violation	23
11	Search or Seizure	805
12	Death or Injury In Custody	756
13	Firearm Discharge – Hits	827
14	Unnecessary Display of Weapon	1445
15	Excessive Force	12494
16	Unlawful Denial of Counsel	10
17	OC Discharge	509
18	Taser Discharge	161
19	Coercion	211
20	Verbal Abuse	1800
21	Domestic Violence	688