**Question 1:**

1. The various assumptions must be made about the data collection process to ensure that valid conclusions about the trends in emissions are reached based on the data. The collection process is assumed that there is minimal bias within the sampling frame. The bias maybe can lead to the under-coverage, an incomplete sampling frame, or because of response and non-response bias.
2. As the number of facilities varies from year to year, so does the sample size, n, of the data. The direct reason for the variability is an increase in the number of facilities will result in an increase in the sample size. Therefore, the data will be more inclusive of facilities with low thresholds. However, in turn the results will become more precise and representative of the actual population size. A decrease in the number of facilities will increase variability and margin of error within the data.
3. The voluntary industrial sources do not exceed a threshold of 50 000 tonnes, it can provide the Government of Alberta with a larger sample size. So researchers can get a wider range of data. However, the participants are voluntary and the possibility for bias is increased. The data may influence the results so it is difficult to take data from voluntary applicants and create a precise conclusion.
4. The change of the mandatory threshold from 100 kt to 50 kt implied that post 2009, the sampling frame increase include many industrial sources. Therefore, there will be an exponential increase in the amounts of data collected. As a result, there is an increase in precision and a decrease in variability within the data. Hence as a result the mean, standard deviation, IQR, and range of the data are affected and become a more accurate representation of the total population. Due to the increase of the threshold, data before 2010 should not be included, as it doesn’t not accurately represent the population.

**Question 2:**

1. **Facilities Reporting GHG Emissions in 2013 and their Total CO2 Emissions**

| **Year** | **n** | **Mean** | **Std. dev.** | **Median** | **Max** | **Q1** | **Q3** | **Sum** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2004 | 98 | 1098314.1 | 2354167.5 | 285429 | 16389317 | 148130 | 794202 | 1.0763478e8 |
| 2005 | 99 | 1079008.5 | 2333753 | 278946 | 15902401 | 137639 | 741866 | 1.0682184e8 |
| 2006 | 103 | 1105351.4 | 2478447.5 | 274873 | 16016481 | 138402 | 701674 | 1.1385119e8 |
| 2007 | 106 | 1060462.7 | 2454238.9 | 257647 | 15390539 | 124717 | 603238 | 1.1240904e8 |
| 2008 | 109 | 1015631.5 | 2299698.9 | 267322 | 14895396 | 132892 | 649437 | 1.1070383e8 |
| 2009 | 93 | 1223739.9 | 2441689.6 | 323268 | 14486722 | 203347 | 977186 | 1.1380781e8 |
| 2010 | 163 | 761269.46 | 2027849 | 149732 | 15787915 | 67704 | 437967 | 1.2408692e8 |
| 2011 | 165 | 756198.77 | 1891679.5 | 171351 | 13029755 | 66636 | 445645 | 1.247728e8 |
| 2012 | 165 | 773399.22 | 1863689 | 151997 | 12530676 | 68877 | 459660 | 1.2761087e8 |
| 2013 | 167 | 794341.38 | 1906664.2 | 165157 | 12548723 | 70983 | 493288 | 1.3265501e8 |

* Total number of facilities: 167
* Total emissions: 132655010 tonnes
* Mean: 794341.38
* Standard Deviation: 1906664.2
* Maximum Value: 12548723
* Q1: 70983
* Q2/ Median: 165157
* Q3: 493288

1. **Summary Statistics for Total (tonnes CO2 eq):**

**Where: “Total (tonnes CO2 eq)” < 50 000**

| **Year** | **n** | **Mean** | **Std. dev.** | **Median** | **Max** | **Q1** | **Q3** | **Sum** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2005 | 5 | 16751.8 | 17203.63 | 15375 | 44887 | 4574 | 17701 | 83759 |
| 2006 | 5 | 18128 | 15213.584 | 17975 | 33744 | 4552 | 32930 | 90640 |
| 2007 | 7 | 11569 | 15982.538 | 2579 | 44451 | 1156 | 17553 | 80983 |
| 2008 | 8 | 12955.125 | 14312.048 | 7637 | 38109 | 1269.5 | 23859.5 | 103641 |
| 2009 | 1 | 7134 | NaN | 7134 | 7134 | 7134 | 7134 | 7134 |
| 2010 | 9 | 24560.222 | 15254.062 | 25182 | 47959 | 10806 | 34962 | 221042 |
| 2011 | 19 | 24316.316 | 18078.35 | 23378 | 49660 | 8066 | 40294 | 462010 |
| 2012 | 20 | 22574.5 | 18402.658 | 18886.5 | 48586 | 4424 | 40313 | 451490 |
| 2013 | 20 | 24488.1 | 18373.638 | 18975 | 48102 | 6659.5 | 42591.5 | 489762 |

* In 2013, 20 facilities reported their emissions voluntarily. The threshold of each of these facilities was below 50000 tonnes. Out of the 167 total facilities, those 20 voluntary facilities produced approximately 489762 tonnes out of the total 132655010 tonnes. Therefore equalling 0.03691997762 % of the total reported greenhouse gas emissions.

1. The facility that emitted the most CO2 emissions in 2013 was the Mildred Lake and Aurora North Plant Sites. It’s total emissions equaled 12548723 tonnes. The 110 lowest CO2 emitters of 2013 had a emission total of 12551204 tonnes.

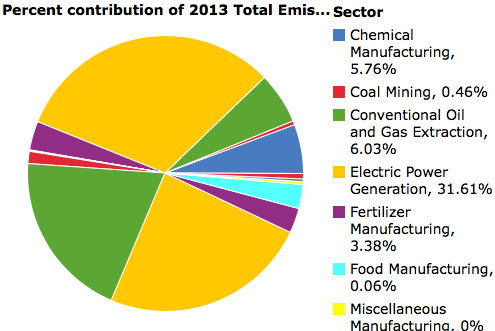
**Summary statistics for Total (tonnes CO2 eq):**  
Where: "Total (tonnes CO2 eq)" < 300000  
Group by: Year

| **Year** | **n** | **Sum** |
| --- | --- | --- |
| 2004 | 50 | 8471388 |
| 2005 | 52 | 8147572 |
| 2006 | 58 | 9500653 |
| 2007 | 60 | 8899148 |
| 2008 | 62 | 9532700 |
| 2009 | 43 | 8012586 |
| 2010 | 108 | 11852088 |
| 2011 | 107 | 11515708 |
| 2012 | 111 | 12943588 |
| 2013 | 110 | 12551204 |

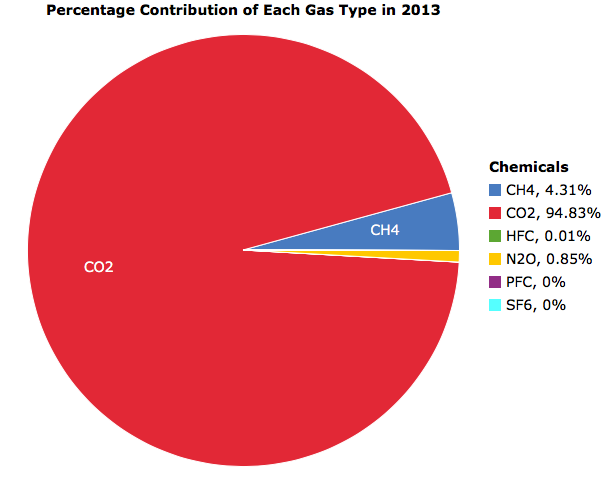
1. The 95th percentile of the total CO2 emissions in 2013 was 4538655 tonnes. The percentage contribution of the remaining 5% of the facilities was 10589 956 tonnes. Meaning the top 5% of facilities contributed approximately 7.983080322%.
2. **Summary statistics for Total (tonnes CO2 eq):**  
    Group by: Sector

| **Sector** | **Sum** |
| --- | --- |
| Chemical Manufacturing | 7647359 |
| Coal Mining | 613974 |
| Conventional Oil and Gas Extraction | 8005222 |
| Electric Power Generation | 41932714 |
| Fertilizer Manufacturing | 4486525 |
| Food Manufacturing | 82665 |
| Miscellaneous Manufacturing | 2559 |
| Natural Gas Distribution | 140260 |
| Non-Metallic Mineral Product Manufacturing | 1892374 |
| Oil Sands In Situ Extraction | 26160060 |
| Oil Sands Mining and Upgrading | 32390594 |
| Petroleum and Coal Products Manufacturing | 3876739 |
| Pipeline Transportation | 3735903 |
| Primary Metal Manufacturing | 449187 |
| Waste Treatment and Disposal | 418145 |
| Wood Product Manufacturing | 820731 |

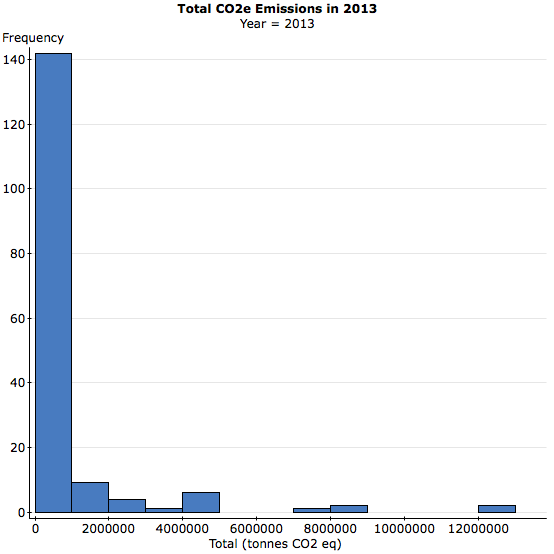
* The Electric Power Generation was the sector with the greatest CO2 emissions. Electric Power Generation contributed 41932714 tonnes. It’s percent contribution was 31.61%.



| Column | Number of Facilities | Sum (tonnes CO2 eq) |
| --- | --- | --- |
| CH4 (tones CO2 eq) | 167 | 5714418 |
| CO2 (tonnes CO2 eq) | 167 | 1.2579956e8 |
| N2O (tonnes CO2 eq) | 167 | 1130662 |
| HFC (tonnes CO2 eq) | 167 | 8579 |
| PFC (tonnes CO2 eq) | 167 | 88 |
| SF6 (tonnes CO2 eq) | 167 | 1703 |



* PFC and SF6 contributed 0% to the total reported GHG emissions. HFC contributed 0.01 % and N2O contributed 0.85%. While CH4 contributed 4.31%. The gas in which contributed the most to the total of GHG emissions was CO2 with 94.83%. Together the six gases make up 100% of the total GHG emissions.



* The tail of the histogram is on the right-hand side, the histogram is right/positively skewed. Besides the tail on the right-hand side we also know that the histogram is positively skewed because the mean is greater than the median and Q3-Q2 is greater than Q2-Q1.
* In 2013, greenhouse gas emissions were collected from 167 industrial sources according to the *Alberta Specified Gas Reporting Regulation* program. Twenty out of the 167 (whose emission thresholds were below 50 000 tonnes) industrial sources voluntarily submitted their emissions. The total 132655010 tonnes of CO2 emissions, were compromised of six gases; CO2 (94.83%), CH4 (4.31%), N20(0.85%), HFC (0.01%), PFC (0.00%), AND SF6 (0.00%). It may be concluded that the total emissions of the majority of facilities have a emission rate equal to or less than 1000000 tonnes. Any values below 5000000 tonnes are rare and the greatest values (above 12000000) are outliers.