“Tate holds the national collection of British art from 1500 to the present day and international modern and contemporary art. The collection embraces all media, from painting, drawing, sculpture and prints to photography, video and film, installation and performance.” The availability of data regarding their collection can provide insights for future acquisitions and raising funds. Additionally, art collectors may find that the data can provide suggestions for future investments.

The data sets used for this project include a set of 3,500 artists with basic bibliographic information, and a set of 70,000 art pieces, which includes titles, artists, date created, date acquired, credit lines, and subject keywords. The credit line property intrigued me the most. To use this feature I needed to parse the text into categories, and clean some mis-categorized pieces. I first looked at trends from the1875 to 2013 to see how the art pieces were obtained. While the number of pieces left to the museum through wills remains steady, there are spikes in the categories where people give pieces to the museum in lieu of paying taxes, where people present pieces in the museum, and large spikes of purchases by the museum. Further exploration into what motivated these spikes will be quite interesting. For example if there was a certain artist availability, change in tax code, or monetary gifts. Looking at the spike in presented in 1975, there were 295 unique artists, and 8% of the pieces were by Moore, Henry, OM, CH. He actually was a trustee of the Tate and had proposed a wing devoted to his sculptures. In 1997 there were a large number of pieces purchased. There were 164 unique artists and 16% of the pieces were by Flaxman, John. But it was actually part of a large collection purchased in 1997. The Oppé Collection was one of the last major collections of British watercolours and drawings to have stayed in private hands. The collection has long been regarded as being of national importance and was acquired by Tate in 1996.

Another factor I wanted to consider was the relative of a piece at the time it was acquired. While most pieces are acquired when they are relatively young (< 20 years), there are a few years when many older pieces, possibly a collection of a single artist, are acquired. Further investigation of these results may benefit collectors who may want to invest in pieces. One feature that may be useful for investors is to see the types of pieces the museum has invested in the past. Future inquiries in the subject matter of pieces may be the most interesting and challenging step. It would be interesting to combine this data with multiple art museum collections.

My last bit of exploration was to look into the birth country of the artists and compare it to the time the art pieces were created. This required mapping the artist information to the artwork data set. While the majority of the pieces are from Europe, there pieces from the different continents. Most pieces are from North America, but the second plot shows when there may have been an art movement in another continent. For example, there are numerous pieces from South America from 1976 and from Asia from 2007. Examining this more closely:

**South America 1976 [2009]**

Ferrari, Leon - His art often dealt with the subject of power and religionFrom Argentina and exiled in 1976 from Argentina to Brazil bc military dictatorship. 2004 controversy forced to close following intervention by Pope Francis (then Archbishop of Buenos Aires) and a subsequent court order. Protests and government action allowed the exhibition to reopen

**Asia 2007 [2008]**

Zaatari, Akram His work is largely based on collecting, studying and archiving the photographic history of the Arab World. 2007 - First solo exhibition

Who am I?

My opinion on a career in health sciences has changed quite a lot since high school. As a freshmen in high school I decided I did not want to be a part of the medical field. I did not want to smell or look at somebody else’s pee nor did I want to dissect pigs. Luckily, my other interest in solving environmental issues could make a positive difference in the world. I was really interested in understanding the problems, but wanted to be more a part of the solution. At first I got involved with the biological side by studying how plants and bacteria could clean up toxic soil, but in college, I really enjoyed the intellectual challenge of studying environmental fluid mechanics and turbulence. I greatly enjoying the research experiences I’ve done, but there were certain aspects lacking in a purely science career. And I know when I love what I do I become obsessive until I’m really good at what I’m doing. I decided to switch to a career in health data science because I want

1. Unfold stories from data and relate it to the big picture!
2. Tackle pressing problems that will benefit society positively
3. A productive, collaborative environment – I love working with other enthusiastic people who are cooperative and like to share ideas
4. A challenge – I am a scholar at heart and I don’t want to ever stop learning
5. I can use the skills I like to hone – computational, mathematical, and analytical skills for a field I do care about

**Questions**

1. What are the next steps?
2. What are the top things to learn before the fellowship?
3. Have I said anything in this interview or given you any other reason to doubt that I am a good fit for the fellowship?

**How maps packed with data help scientists fight malaria**

Scientists have begun using temperatures, rainfall patterns and other data to better target areas most at risk. Could map cases of ebola and distribution of fruit bats. http://www.pbs.org/newshour/bb/maps-packed-data-help-scientists-fight-malaria/

1. Broad Institute
   1. Genome Analysis Toolkit (GATK)
      1. Analyze genetic information and detect genetic variants associated with medical conditions for Cancer, Neurodegenerative Disorders and Cardiovascular Disease
      2. Partnership with Bayer to search for genetic biomarkers of patient risk or suitability to particular drugs that treat cardiovascular conditions like heart disease
      3. **Sequencing of Ebola virus mutations - The Ebola strains responsible for the current outbreak likely have a common ancestor, dating back to the very first recorded outbreak in 1976. The lineage responsible for the current outbreak diverged within the last ten years and spread from Guinea to Sierra Leone by 12 people who had attended the same funeral.**
         1. **Speed up our understanding of this epidemic and support global efforts to contain it.**
2. Novartis Institutes for BioMedical Research
   1. Genomic data – specifically Next Generation Sequencing (NGS) data
      1. Genetically identify patients who will respond positively to particular therapies
      2. Personalized cancer treatment
   2. Accelerate drug discovery
   3. Oncology, ophthalmology, and cardiovascular and metabolic related diseases
   4. **Partnership to use Google contact lens technology to monitor glucose levels in tears of people with diabetes and/or to provide accommodative vision correction**
3. Memorial Sloan Kettering Cancer Center
   1. Developing innovative approaches to improve cancer research by marrying the latest methods in predictive modeling and data mining with traditional biostastistics and epidemiological methods
   2. End of Life Care: There is a significant amount of literature studying where people spend their final days: in a hospital, in a hospice, or at home. A non-trivial portion of our patients pass away in the inpatient hospital. Are we providing them the most effective care there? What value are we providing them, and are there any alternatives?
   3. Readmissions: We track rates at which patients return to the hospital for followup care. Recently, our rates have been declining, even in the absence of particular initiatives or improvement projects. Can we find out what the root cause is?
   4. Quality of Care: How do we compare, in terms of outcomes and quality, to other hospitals? Can we adjust for metrics such as severity of illness and case mix?
4. Biogen Idec
   1. Gave Fitbit activity monitors to people with MS to assess their reliability as data-gathering tools. And it is in the middle of developing an iPad app with Cleveland Clinic to better track disease progression. Now, Biogen plans to work with Google to gather and analyze data on the lives of people with MS.
   2. Impact of new drugs
   3. **Alzhiemer’s**
      1. **Treating Alzheimer’s in most patients will require a combination of therapies**
      2. **Screened patients and found that 40% didn’t have it despite signs of dementia - looked for amyloid deposits in the brain**
      3. **Possibly use drug to prevent/treat disease**
5. MassBio
   1. Not sure…It supports biotech companies
6. Infinity Pharmaceuticals
   1. Difficult-to-treat diseases: hematologic malignancies, or blood cancers. We are focused on advancing duvelisib for the treatment of indolent non-Hodgkin lymphoma (iNHL) and chronic lymphocytic leukemia (CLL).
   2. Aggregate research data in research and development efforts to more quickly predict drug development outcomes, thus reducing the costly paths to take new drugs to market.
   3. Predictive modeling can shave three to five years off the approximately 13 years it can take to bring a new compound to market
   4. Design clinical trials and target patients for clinical trials. Moreover, analytics can be vital in monitoring and mining clinical trial data and patient records to identify negative effects or benefits from the drug’s use
7. Invitae
   1. Specializing in genetic diagnostics for hereditary disorders
   2. Utilization of genomic data in order to improve health, catalyze scientific and medical research, and contribute to improved outcomes for everyone
   3. **Multi-gene hereditary cancer tests - those without BRCA1/2 mutation advised care would be different b/c of mutations in other cancer-risk genes. Previously it was just on family history.**
8. Optum
   1. High-quality integrated health care data, covering millions of lives with de-identified claims and clinical data that is cleansed, tested and refreshed regularly. This gives researchers an expansive view into each research scenario and allows for unique insights into population health.