Big Data Analytics for Municipal Waste Management

Andres Castro Benavides Indiana University 107 S. Indiana Avenue Bloomington, Indiana 43017-6221 acastrob@iu.edu Mani Kumar Kagita Indiana University 107 S. Indiana Avenue Bloomington, Indiana 43017-6221 mkagita@iu.edu

ABSTRACT

Waste management is becoming a greater concern for cities and municipalities around the world because of the continual increase in population and waste. Big data analysis has the potential to not only help assess the current waste management strategies but also provide information that can be used to optimize the systems used in various institutions, local governments, and companies, among others.

KEYWORDS

Waste Management, Big Data, Local Government, HID305, HID319

1 INTRODUCTION

In the current fast-paced society, production of goods continues to increase, and new distribution chains continuously change. The generation of waste and the deprecated goods -from now on referred to as solid waste- has increased over the past ten years, rising from approximately 0.64 kg per person per day of solid waste to approximately 1.2 kg per person per day. Projections estimate that this number is expected to increase to about 1.42 kg by 2025. [15] This continual change and increase make waste management a more complex and more intensive endeavor.

While the quantity of waste itself can and should be reduced by conscious use and a discipline in recycling and reusing items, there is much that can be done to reduce waste in the actual collection and disposal by optimizing waste management systems.

Because of this, different local governments and organizations have recognized the need to develop more elaborate regulations to control the different features, segments, processes involved in waste disposal from the moment the material is discarded by the consumer to the moment the material reaches its ultimate destination, such as a recycling plant or landfill. This set of systematic regulations is called solid waste management. Solid waste management has changed over time. What used to be systems designed and implemented based on local needs and convenient disposal as moved into extensively researched and implemented management systems that consider complex multivariable and dynamic sources of data. [1].

Currently, data used in waste management is collected from many sources and varies depending on the types of solid waste and the rate of disposal. Because of the diversity of sources available (including types of waste, weight collected, the location of collection and disposal, etc.), the quality of the data must be continually monitored and assessed. [6]. New management systems that seek to optimize waste management must collect large volumes of data from each data source on a daily basis. Multivariate data analysis methods provide exploratory data analysis, classification and parameter prediction using this data. [4]

2 MUNICIPAL SOLID WASTE MANAGEMENT

There are significant differences between the general composition of the waste generated in rural areas and the waste produced in urban areas. The waste produced in the later is profoundly influenced by the culture and the practices of our modern society. This distinction means that there will inevitably need to be differentiation in waste management practices and systems based on each communities needs, commerce, economy, and practices. [6]

Municipal Solid Waste (MSW), commonly known as garbage or trash, consists of items that residential, commercial, institutional and industrial sites generate. This is much different from the type of waste produced, for example, in an agricultural community. Urban, or municipal, waste can be the result of everyday activities, such as leftover food, plastic bottles, packaging material, wooden furniture, electrical and electronic appliances, glass, medical waste, cardboards, waste tires, office wastes, consumer goods, among others.

The amounts of solid waste and its composition vary depending on the country, place, and activity performed at the site where the waste is generated [6]. For this reason, every process related to waste management-transportation, storing and final disposition, among others- must be engineered and tailored to fit the specific needs of community, organization, or local institution. The data collected for analysis will also be specific to the needs and realities of each individual entity, which is why waste management plans are not universal or easily applied from one community to another.

The ways of handling waste have also changed over the years, affecting the best ways to manage waste. According to EPA statistics from 2014, Americans have generated about 258 million tons of MSW of which more than 89 million tons is recycled and composted. The equivalent to a 34.6% recycling rate compared to 6.4% in 1960. Americans have also used the energy production process to combust approximately 33 million tons of waste, while as much as 136 million tons of waste ended up in landfills during the same year. [9]. Because of this, data on recyclable material, dividing waste and recyclable material, and where materials are sent or how they are disposed of has also become significant. Figures 1,2,3 and 4 [9] represents MSW generation rates, recycling, composition rates and Total MSW generation between 1960 and 2014 [9].

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

In waste management, decision-makers are and will continue to be forced to make choices. When they develop or implement plans, they have a choice as to what information they will or will not consider for analysis. They have to choose what factors are influential in their jurisdiction. They make choices about routes, resources, and all the details between pick up to disposal. These choices can be classified as fortuitous, good, or optimal. [1].

Fortuitous decision-making has no scientific base; the person who is in charge of making decisions must always try to solve the problem with little or no research or data. On the other hand, good decision-making is primarily based on experience, comparison of elements and trial and error. Optimal decision making, however, requires understanding and analysis of techniques and technologies provided by other fields. [1]

This is where Big Data comes in.

3 BIG DATA AND WASTE MANAGEMENT

Big data has the capacity to facilitate analysis of information so that better decisions can be made.

In Big Data, the expert in charge takes many data sets coming from diverse and dynamic data sources and applies technologies to analyze these data sets. Authors of papers and books like "The Fourth Paradigm" state that big data exploration works to find patterns in data by analyzing the trends and outliers found in the datasets mentioned above, to generate knowledge. [14]

Characteristics (such as the large volume of data generated in waste management, how dynamically the data is generated, and the variety of formats in which the data comes) turn the task of producing knowledge an ideal task for Big Data. Scientists can interpret the findings of Big Data in a way that allows individuals and institutions involved in waste management, to make optimal decisions. [21]

4 SOLUTIONS FOR EFFECTIVE WASTE MANAGEMENT

4.1 Examples of Big data and waste management

In Manchester, England, the Greater Manchester Waste Disposal Authority, England's most significant waste management institution, has started to use Big Data to better orchestrate the waste management services they provide. To get the most out of their Big Data approach, the Greater Manchester Waste Disposal Authority is working in collaboration with the research done by the University of Manchester. Together, they are creating environmentally sustainable solutions for Manchester, and are finding optimal solutions to the 1.1 million tonnes of waste that Manchester produces every year [19].

Another example of a local government implementing Big Data Solutions is the city of Songdo, South Korea. In said city, every citizen needs to use a chip card while disposing of their garbage. Data collected from these chip cards is being used to analyzing on the quantity of disposed waste and their locations. Each trash bin is incorporated with sensors to provide the height of the garbage accumulated in the reciprocal, temperature, and air pollution levels. These multiple parameters help municipal authorities forecast ideal

times to collect the trash and optimize the routes to save time and expenses [19].

Researchers in Ethiopia are combining socioeconomic data alongside geographic data in order to get a clearer understanding of the patterns of how household waste is being collected and distributed. This study helped local authorities to better manage waste practices in urban areas [19].

A group of researchers from the University of Stockholm is using Big Data to identify how to optimize waste management routes in their city. By using a wide variety of data sets collected from various sources, roughly around half a million entries including trash bin locations, weights, and truck routes, researchers have developed waste generation maps of Stockholm. This research has helped reveal various inefficiencies in the current waste management system and will be integral in helping them improve their local waste management [19]. Figure 5 [18] represents heatmap of all waste aggregated per zip code in Stockholm.

[Figure 5 about here.]

4.1.1 Vehicle Routing Problem. Vehicle route optimization is one of the primary concerns in waste management. It is termed as Vehicle Routing Problem (VRP) [7]. While routes may be designed based on geography, merely looking at a map and designing what seems to be practical movement through the city, this is underestimating the extent of the factors involved. There are multiple factors that either directly or indirectly affect waste collection that can also be analyzed and considered while designing routes. Common known factors that influence vehicle routing are the type of vehicle, vehicle capacity, number of collection stops, volume per capita and the route length. Figure 6 represents Vehicle Routing Problem [20] when one dump location is placed between different collection routes.

[Figure 6 about here.]

Two of the most basic VRPs are the Travelling Salesman Problem (TSP) and the Chinese Postman Problem (CPP) [2]. However, when too many constraints and attributes are considered, both of the TSP and CPP become more difficult problems to solve. Many researchers have studied and published articles on waste management vehicle routing problemVRPs, and yet it is a persistent problem. Mathematical models need to be developed to provide city administrators with tools to make effective long-and short-term decisions relating to their municipal disposal system [3].

In addition to the common known factors causing VRP problems, indirect attributes must also be taken into consideration. Time of day, traffic, weather patterns, energy prices, demand fluctuations, vehicle health, dump site inventory also have potential to affect route efficiency.

A research team at OSI came up with a better solution for solving VRP problems using Big Data technologies. Mixed Integer Programming (MIP) formulation was designed to interact with millions of attributes in live environments providing real-time decisions to optimize the VRP. Big Data technologies are being used to enable prediction of travel times and forecast and address demand on a tactical timeline. This approach has helped improve VRP forecasting between 5% to 10% [12]. Any improvement, even less than 5% created on VRP is a significant improvement [13].

Scientists and engineers have designed a system that communicates between waste reciprocals, waste collection vehicles and a central system using sensors that measure how full containers are and can send real-time data. [10].

The data collected can be used to optimize routes and space in the waste collection vehicles. Sensors can identify which containers do not need to be collected, allowing them to shorten routes. In addition, sensors can measure the remaining capacity of collection vehicles, allowing them to extend routes when they still have cargo space, ensuring that they return to the disposal center only when they are carrying a full load. By using these sensors, fewer collection vehicles were needed. It was estimated that in three years, the expense of purchasing and implementing this system would be recovered [18].

In addition to the real-time data collection that helps optimize collection and routes, remote self-diagnostics are also helping optimize vehicle use. By being able to monitor the vehicles use and health, maintenance and repair are more efficiently managed. Through this system, managers are made aware of parts that need to be ordered in advance, limiting the time that the vehicles may be out of service. Handheld devices have been developed for service verification, further helping minimize external providers or assessments, optimizing asset management and costs [11].

[18].

4.1.2 Problem of Landfill Disposal. A municipal solid waste landfill (MSWLF) is a individually separated area of land or a trench where the household waste is collected and stored. These landfills are designed to store municipal solid waste as well as other wastes like construction and industrial waste. These landfills can be openpits or below ground refuse chambers. In recent years, its becoming more expensive to operate and maintain these landfills as to protect the environment from escaping liquid pollutants and causing water and surface contamination. These liquid pollutants are commonly called as leachate. It forms when water originating from rainfall or groundwater dissolves gradually through a porous surface and dissolves the chemicals from the refuse especially if the protective layer of the landfill allows liquid or gases to pass through it.

Leachate contamination problems became more problematic in older landfill sites when they are typically lacked in barriers above or below the landfills which causing pollution and diseases to the citizens residing in areas near to these landfills. In modern days, Governments across the world are in one motive to prevent waste disposal and reduce landfill deposits. Steps are taken to reduce number of landfill areas and extending landfill capacity at current sites. Figure 7 displays the number of landfills in the US, 2009.

[Figure 7 about here.]

Landfill disposal, in itself, has its own set of complex factors including measurement and control of gases, management of wastewater, and precipitation patterns affecting water volume in the geological basin.

Identifying these factors creates an opportunity to develop sensors that could collect meaningful data that would be an asset to optimizing landfill disposal. Sensors could measure gas immission, water volume in disposed waste, and precipitaion. Having this information would allow data scientists to develop systems that would identify when disposal would need to be varied, how water

treatment plants could efficiently manage runoff and basin water and communicate with water treatment plants to use more or less of their mechanical resources based on real-time needs.

Manual calculation would be time-consuming and physical observation opens space for inefficiency and delay. In order to compute these calculations promptly so that the information could contribute to optimization, it would be ideal to develop a computer program to compute the mathematical operations. This would contribute significantly to optimization of water treatment, for example, in landfill disposal [1]

Waste management is already in the process of transforming from using older methodologies to modern Big Data technologies. Big Data has a lot more potential to eradicate most of the problems faced by government in waste disposal. Most of the government organizations are taking a leap to minimize disposal of waste and to achieve "zero disposal" goals [17]. A program in government of District of Columbia called "Zero Waste DC" is being initiated to develop and provide resources that will help its residents move towards zero waste. Its main motive is to divert 80% of MSW to recycling or source of energy where the remaining 20% non recoverable waste will be sent to landfills. By collecting vast amounts of data from all sources, they will analyze it to implement costeffective strategies for converting waste to re-usable resources, improving environmental conditions, taking measures on human health, reducing greenhouse gas emissions and conservation of natural resources. Obtaining data from the sensors on contamination of air with harmful gases, water purification levels from the nearby lakes, physical properties as well as chemical properties of the soil, Big Data will help to analyze this data and provide solutions to prevent them from making situations worst. [17].

When its considered Big Data as a volume, huge amounts of data is being collected daily from different sources like digital meters, sensors, social media to avoid waste disposals. It is how this data is being turned into knowledge to apply into real-world problems. Cloud based solutions and using IoT sensors becoming more insight technologies for smart cities and municipal waste management authorities to optimize the waste and their recycle operations. Big Data will help in lowering of contamination on the waste and improve consumption habits of residents in urban areas.

Reading information from the sensors placed in recycle bins, it is easier to figure out what type of waste is being disposed in the trash bins and by which resident. If there are hazardous materials located, authorities can detect them before even being collected by the trash collectors. This will help to educate corresponding resident on the importance of waste and their contamination results. Big Data Analysis encourage citizens to understand about food scraps so that it can be reused as fertilizer for many of the organic foods. It is termed as "farm to farm" practice according to Robert Reed, Evangelist at Recology [16].

Using Big Data technology to collect real-time data thought the waste disposal process especially for food wastage allows visibility to organic waste. This shows focus on identifying inefficiencies in food management process which in turn helping them to initial improvement in process and to create immediate effects. Optimizing waste management is a key to find operational efficiencies and to support environmental directives. It enables business to

make effective decisions about purchasing or production of organic waste [5].

Environmental Protection Agency US also using benefits of Big Data in recycling of waste. Big Data Analysis helps in identifying quantities of solid waste being recycled and assessing the residents on the benefits they receive by recycling achievements. MSW Characterization Report shows the data collected on generation and disposal of waste in US past 30 years. This data is being analyzed to measure waste reduction and plan recycling programs across the country [8].

4.2 Statistics and Waste Management

Statistics help develop mathematical models to predict behavior based on the data collected. In waste management, these statistics can be used by Big Data professionals to draw conclusions and foresee possible outcomes as related to effective policy making and actual waste production, collection, and disposal patterns.

There are many data analysis methods that are used when studying waste management, but the two most popular are PCA and PLS1. [4]

Lingo is a mathematical modeling language designed particularly for formulating and solving a wide variety of optimization problems including linear programming. Lingo optimization software uses the branch and bound methods to solve problems of this type. [1].

These resources are valuable to the continual analysis of data being collected, continually helping in waste management optimization.

4.3 GIS Analytics

When it comes to Geographical Information Systems (From now on GIS), there are multiple software and hardware options in the market. From paid software like ArcGIS to Open and free software like GVSIG, some solutions can help interpret large data sets, apply statistics and algorithms of different kinds and display them in a way that refers to a geographical space.

Two optimal routing algorithms have been used to calculate routes for waste collection; Solomon's insertion algorithm and a clustering algorithm. Data was used to help create more efficient routes by minimizing driving distance while taking into consideration factors like lunch breaks (time windows) that affected the distribution of human resources and time/vehicle management. By adding vehicle depots, by rerouting, and by sharing various routes, they predicted that they would be able to shorten vehicle expenditure by as much as 10,000 km.

An additional routing algorithm was used to present information on the environmental significance for optimizing waste collection routes. The study assessed the use of roll-on/roll-off containers while also considering schedules and lunch breaks. Through the utilization of an adaptive large neighborhood search algorithm, alongside a clustering method, a residential waste collection was analyzed, taking into consideration actual pick up points and employee lunch breaks. Using this information, time windows and starting conditions were adapted alongside the routes to reduce the total distance by as much as forty-five percent. [18].

5 CONCLUSION

Waste management has been a growing concern and will continue to be an important area for optimization as both consumer waste and population increase. As institutions, governments, and individuals look to assess and optimize resources, minimize cost and create less of an environmental impact of waste management, Big Data has the potential to continue to help provide the information needed for future advancement.

There are various tools being used to optimize the different waste management practices, and there is space to develop additional tools to continue to the information available to decision makers.

One of the main reasons to use Big data in Municipal Waste Management is to provide local governments with tools that would facilitate the implementation of systems to more efficiently manage how much, where and the growth rate of the material that the community dispossess. Optimizing the waste management systems can help minimize the environmental impact of the community's actions, reduce pollution, increase rates of recycled materials, optimize routes to reduce time and expenses, among others. By using Big data, local governments can also track the number and quantity or weight of disposals at different locations; this information can be mapped to reveal the locations of the most significant waste generators. This information will help entities develop specific strategies to reduce waste and help implement permanent solutions for better environments. Waste Management is not only government issue. Citizens should take the initiative and educate others on how to recycle and reduce waste. With the help of the collected data, governments can notify the different entities (individuals, communities, companies and other organizations) a equip them through education and awareness, and also share valuable information about the importance of waste management through different media, such as mobile phones, email, among others. Local governments have just started to adopt Big Data technologies for solving problems involved in MSW, but there is plenty of room for growth and further use and development. By using Big Data Analytics, large amounts of data sets pertaining to specific communities waste, routes, and disposal can be used to identify trends and patterns that could highlight opportunities for improvement. Big Data can play a significant role in managing cities more efficiently, benifiting not only those managing the systems, but the commuties in which they are implemented as well.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions in writing this paper.

REFERENCES

- [1] Mohsen Akbarpour Shirazi, Reza Samieifard, Mohammad Ali Abduli, and Babak Omidvar. 2016. Mathematical modeling in municipal solid waste management: case study of Tehran. Journal of Environmental Health Science and Engineering 14, 1 (18 May 2016), 8. https://doi.org/10.1186/s40201-016-0250-2
- [2] Jeroen Belin, Liesje De Boeck, and Jonas Van Ackere. 2012. Municipal Solid Waste Collection and Management Problems: A Literature Review. HUB RESEARCH PAPERS 2011/34 ECONOMICS & MANAGEMENT 48, 34 (11 2012), 1–5.
- [3] V. N. Bhat. 1996. A model for the optimal allocation of trucks for solid waste management. A model for the optimal allocation of trucks for solid waste management. 14 (1996), 87–96.
- [4] K. Bofkhm, E. Smidt, and J. Tintner. 2013. Application of Multivariate Data Analyses in Waste Management. In Multivariate Analysis in Management, Engineering

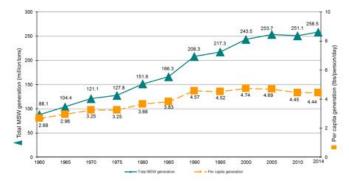
- and the Sciences, Leandro Valim de Freitas and Ana Paula Barbosa Rodrigues de Freitas (Eds.). InTech, Rijeka, Chapter 02, 15–16. https://doi.org/10.5772/53975
- [5] Frank E. Celli. 2016. Data analytics: The most effective approach for a zero waste solution. WasteDive. (09 2016). http://www.wastedive.com/ news/data-analytics-the-most-effective-approach-for-a-zero-waste-solution/ 425424/
- [6] R. Chandrappa and J. Brown. 2012. Solid Waste Management: Principles and Practice. Springer Berlin Heidelberg, Berlin. 47–63 pages. https://books.google. com/books?id=kUOwuAAACAAJ
- [7] G. B. Dantzig and J. H. Ramser. 1959. The Truck Dispatching Problem. Management Science 6, 1 (10 1959), 80–91.
- [8] EPA. 2007. METHODOLOGY FOR ESTIMATING MUNICIPAL SOLID WASTE RE-CYCLING BENEFITS. epa.gov. (11 2007). https://www.epa.gov/sites/production/ files/2015-09/documents/06benefits.pdf
- [9] EPA. 2014. Advancing Sustainable Materials Management: Facts and Figures.
 U.S. Environmental Protection Agency. (2014). https://www.epa.gov/smm/advancing-sustainable-materials-management-facts-and-figures#Materials
- [10] Maurizio Faccio, Alessandro Persona, and Giorgia Zanin. 2011. Waste collection multi objective model with real time traceability data. Waste management (New York, N.Y.) 31, 12 (08 2011), 2391–405. https://www.ncbi.nlm.nih.gov/pubmed/ 21821406
- [11] Megan Greenwalt. 2017. What the Growth of Big Data Means for Waste & Recycling. FLEETS & TECHNOLOGY. (03 2017). http://www.waste360.com/ fleets-technology/what-growth-big-data-means-waste-recycling
- [12] Vijay Hanagandi. 2013. A New Paradigm to Solving Vehicle Routing Problems. (09 2013). https://osiblogdotcom.wordpress.com/2013/09/23/ a-new-paradigm-to-solving-vehicle-routing-problems/
- [13] Geir Hasle, Knut-Andreas Lie, and Ewald Quak. 2007. Geometric modelling, numerical simulation, and optimization: Applied mathematics at SINTEF. Springer, Berlin, Heidelberg, Oslo, Norway.
- [14] A.J.G. Hey, S. Tansley, and K.M. Tolle. 2009. The Fourth Paradigm: Data-intensive Scientific Discovery. Microsoft Research, REDMOND, WASHINGTON. https://books.google.com.my/books?id=oGs_AQAAIAAJ
- [15] Perinaz Hoornweg, Daniel; Bhada-Tata. 2012. A Global Review of Solid Waste Management. Number 15 in Urban Development Series. World Bank, Washington, DC, Urban Development & Local Government Unit World Bank 1818 H Street, NW Washington, DC 20433 USA. https://openknowledge.worldbank.org/handle/ 10986/17388
- [16] James Kobielus. 2012. Reuse, Recycle, Compost: A New Level of Insight into Garbage. Robert Reed, Evangelist at Recology, zerowasteIBM. (o6 2012). http://www.ibmbigdatahub.com/blog/ reuse-recycle-compost-new-level-insight-garbage
- [17] Cole Rosengren. 2017. San Francisco expands recycling list, shrinks refuse carts on 'zero waste' crusade. WasteDIVE. (10 2017). http://www.wastedive.com/news/ san-francisco-expands-recycling-list-shrinks-refuse-carts-on-zero-waste/ 506700/
- [18] Hossein Shahrokni, Bram Van der Heijde, David Lazarevic, and Nils Brandt. 2014. Big data GIS analytics towards efficient waste management in Stockholm. In Proceedings of the 2014 conference ICT for Sustainability. Atlantis Press, Proceedings of the 2014 conference ICT for Sustainability, Department of Sustainable Development, Environmental Science and Engineering, Industrial EcologyRoyal Institute of TechnologyStockholm, Sweden, 140–147.
- [19] Mark van Rijmenam. 2016. How Big Data Shapes Urban Waste Management Services in Manchester. techreport. University of Technology, Sydney. https://datafloq.com/read/how-big-data-shapes-urban-waste-management-service/ 662
- [20] wikipedia. 2017. Vehicle routing problem. wikipedia. (2017). https://en.wikipedia. org/wiki/Vehicle_routing_problem
- [21] Vitthal Yenkar and Mahip Bartere. 2014. Review on fiData Mining with Big Datafi. International Journal of Computer Science and Mobile Computing 3, 4 (2014), 97–102.

6 BIBTEX ISSUES

7 ISSUES

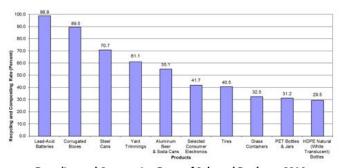
List of Figures

1	MSW Generation Rates in US from 1960 to 2014	7
2	Recycling and Composting Rates of Selected products in US, 2014	7
3	MSW Recycling and Composting Rates in US from 1960 to 2014	7
4	Total MSW generation in US, 2014	8
5	Choropleth of all waste, aggregated per zip code and normalized to population in Stockholm	8
6	Vehicle Routing Problem	9
7	Number of Landfills in the US 1988-2009	9



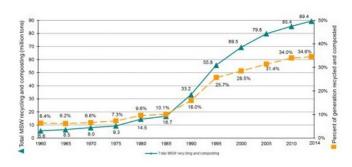
MSW Generation Rates 1960 to 2014

Figure 1: MSW Generation Rates in US from 1960 to 2014



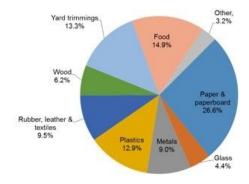
Recycling and Composting Rates of Selected Products, 2014 (does not include combustion with energy recovery

Figure 2: Recycling and Composting Rates of Selected products in US, 2014



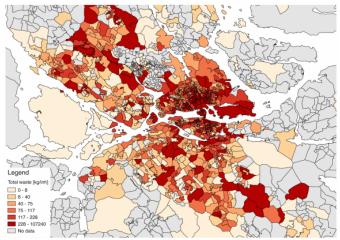
MSW Recycling and Composting Rates, 1960 to 2014

Figure 3: MSW Recycling and Composting Rates in US from 1960 to 2014



Total MSW Generation (by material), 2014. 256 Million Tons (before recycling, composting, or combustion with energy recovery)

Figure 4: Total MSW generation in US, 2014



Choropleth of all waste in the dataset, aggregated per zip code and normalized to population.

Figure 5: Choropleth of all waste, aggregated per zip code and normalized to population in Stockholm

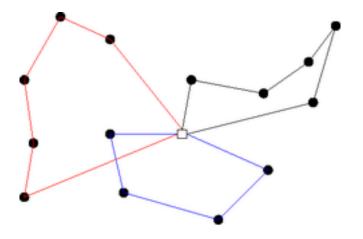


Figure 6: Vehicle Routing Problem

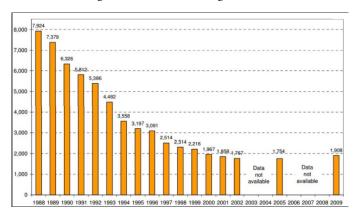


Figure 7: Number of Landfills in the US 1988-2009