ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

**Факультет информатики, математики и компьютерных наук**

**Программа подготовки бакалавров по направлению**  
**09.03.04 Программная инженерия**

*Доронина Оксана Сергеевна*

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

Использование визуализации при анализе экономических данных

|  |  |
| --- | --- |
| Рецензент  доц., канд. техн. наук  А.В. Савченко | Научный руководитель  доц., канд. техн. наук  М. В. Шишкин |

Нижний Новгород, 2016

Оглавление

Введение

Глава 1

Часть 1.1

Введение

Проблема обработки данных наиболее актуальна в наше время. В современном мире существует огромное разнообразие данных и их систематизация, способы представления тоже становятся разнообразнее. Если в старину доступ к данным имел не каждый человек, а только избранные, то теперь каждый человек может найти интересующую его информацию, например, о бюджете страны, или о средствах, потраченных на ремонтные работы, найти данные об изменениях цен на продукты и услуги ЖКХ. Существует много видов обработки данных, но наиболее лучшим является визуализация.

Визуальная информация лучше воспринимается и позволяет быстро и эффективно донести до зрителя собственные мысли и идеи. Физиологически, восприятие визуальной информации является основной для человека. Есть многочисленные исследования, подтверждающие, что:

* 90% информации человек воспринимает через зрение
* 70% сенсорных рецепторов находятся в глазах
* около половины нейронов головного мозга человека задействованы в обработке визуальной информации
* на 19% меньше при работе с визуальными данными используется когнитивная функция мозга, отвечающая за обработку и анализ информации
* на 17% выше производительность человека, работающего с визуальной информацией
* на 4,5% лучше воспоминаются подробные детали визуальной информаци

Именно по ряду вышеуказанных причин, для донесения информации и для ее лучшего усвоения лучше всего использовать визуализацию данных.

Цель данной работы- использование визуализации в анализе экономических данных.

Задачами данной работы являются:

- Поиск подходящих экономических данных;

-Анализ данных;

-Построение интерактивных графиков для выбранных данных

Работа состоит из введения, первая и вторая главы, заключения.  
В первой главе- теоретическая основа данной работы. Во второй практическая часть.

Глава 1.

1.1 Визуализация данных

Визуализация данных — это представление информации при помощи различных графический инструментов, таких как диаграммы, графики, схемы, таблицы.

Важность визуализирования данных состоит в том, что она позволяет наиболее верно и показательно выявить и отобразить информацию, которую содержат данные, потому что на примере изображения легче продемонстрировать то, что в вербальном эквиваленте представляет собой довольно обширный кусок текста.

Благодаря тому, что информационные технологии быстро развиваются и распространяются, появляется большое разнообразие инструментов для создания различных видов визуализаций. Как итог, визуализирование данных стало все чаще использоваться во многих сферах деятельности, прежде всего в научной, медицинской и бизнес-аналитики.

В создании методов оптического представления данных в равной степени принимают участие как IT-специалисты и дизайнеры, так как о разработанного дизайна во многом зависит понятность и «читаемость» визуализации.

С помощью визуализации данных решаются самые разные задачи.

Во-первых, это основной способ для анализа данных на начальных стадиях. Самые элементарные графики позволяют быстро обнаружить взаимосвязи в данных, их закономерности, тенденции, на которые будут опираться аналитики при дальнейшей работе с данными.

Во-вторых, визуализирование данных, зачастую играет немаловажную роль в представлении окончательных результатов анализа. Это могут быть статичные графики, интерактивные визуализации, для самостоятельного исследования данных пользователями; а также инфографика, которая представляет историю, основанную на данных.

Следующая задача визуализации состоит в том, чтобы на одном изображении показать многообразие содержащихся в данных измерений и взаимосвязь между ними, при ограниченном количестве физических соизмерений.

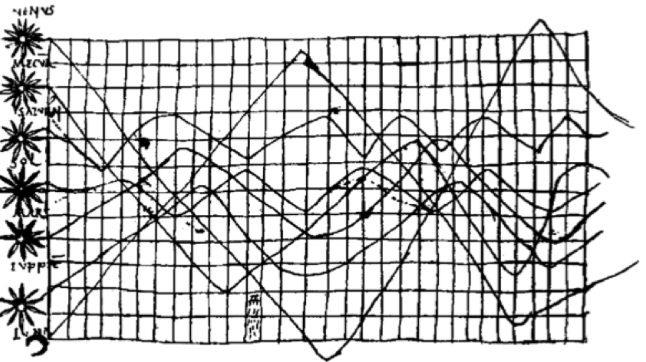
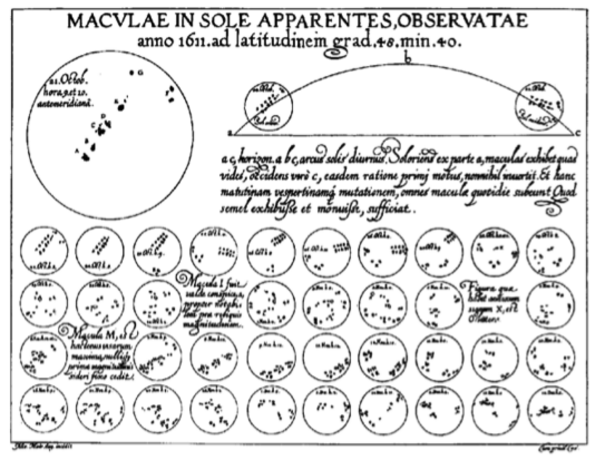
 По данным исследования ученых, люди воспринимают информацию визуально на 95 процентов, а само понятие визуализации данных пришло в современность из далекого прошлого. Наиболее полно историю развития представил автор Майкл Френдли (Michael Friendly) в своем труде Handbook of Data Visualization. Самый древний пример визуализации данных это линейная диаграмма. Первые зачатки визуализации зародились в геометрических диаграммах, в таблицах положения звезд, навигационных картах. Среди ранних изображений информации есть график перемещения небесных тел, где через двумерную систему координат показан график перемещения планет.

Рис. 1.1.1

Данный график был создан до 17 века. Следующий этап развития визуализации данных это 1600- 1699 года, данный период автор в своей книге называет «Измерения и Теории». В семнадцатом веке ученых интересовало, каким способом, возможно оценить, измерить время, скорость, пространство и в основном упор происходил на карты и навигацию. Именно в данном периоде возникли такие понятия как система координат, родилась теория вероятности и демографическая статистика. Ярким примером визуализации данных того времени является работа Кристофера Шайнера (Christopher Scheiner), датированная 1630 годом.  Позднее Эдвард Таффи дал ей определение «маленькие множества», подразумевая под ним повтор одного элемента множество раз для отображения изменений и динамики.

Рис 1.1.2

Следующий этап развития визуализации датируется 1700-1799 годами. Этот период получил название «Новые географические формы», восемнадцатый век это временной отрезок, для которого характерен выход за рамки «очевидного». На картах в то время старались отобразить не только точку географического нахождения, но так же прорисовывать контуры найденных объектов. Тематические карты географических разломов, экономические выкладки и медицинские иллюстрации являются примерами данной эпохи. В то время происходит значительное увеличение информации о политической и экономической обстановке, благодаря чему возникает необходимость в новых формах визуализации для их отображения.

Ниже приведен пример изображения нанесения на карту дополнительных данных.

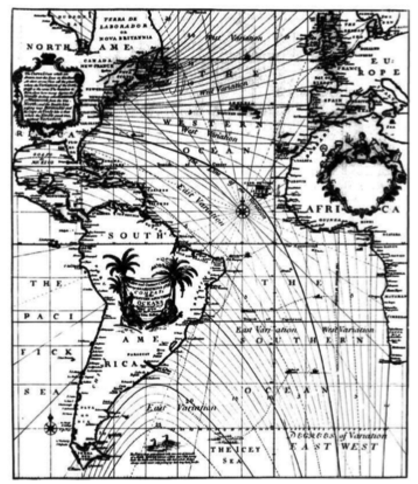
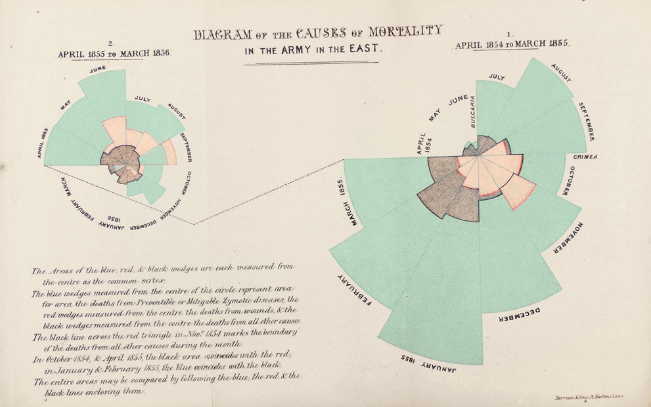


Рис. 1.1.3

Следующий этап развития визуализации датируется 1850 годом и носит название «Начало современной графики» Именно в этот период появляются такие виды графического отображения как круговая, столбчатая, площадная диаграммы. Точкой отсчета для современной графики принято считать карту Лондона, созданную Джоном Сноу, на которой отображены области распространения холеры. На Карте Сноу отмечены кластеры вспышек холеры.

Рис. 1.1.4

Следующее знаменательное событие данного периода произошла в 1858 году, в эти 12 месяцев сестрой милосердия и великобританским деятелем Флоренс Найтингел была изобретена первая круговая диаграмма, которая использовалась во времена Крымской войны, для того, чтобы показать, что большинство солдат погибло от болезни (синий цвет на диаграмме), чем на поле боя( красный цвет на диаграмме) или по другим причинам( черный цвет на диаграмме).

Рис. 1.1.5

Следующий временной отрезок с 1900-1950 года, не является особо значимым для развития визуализации. В это время люди делились на «более визуальных» и «более табличных», например, британский народ считал себя «более табличным». В те времена девизом Британской академии был сбор данных, а не их отображение.

Следующий отрезок во времени стал для развития визуализации эпохой Возрождения, данный период датируется 1950-1975 годами. В то время идеи визуализации стали быстро распространяться среди ученых и литераторов. В 1962 году на свет появилась книга «Будущее анализа данных», автора Джона Тьюки, в которой он отделил математику от статистики. Если первая не признает визуализации, то статистика благодаря отображению данных, обретает более глубокий смысл и более полную форму. Так же это время знаменательно тем, что именно в этом периоде возникает первая интерактивная визуализация.

Примером служит визуализация Ричарда Бейкера, которая будет приведена на рисунке ниже

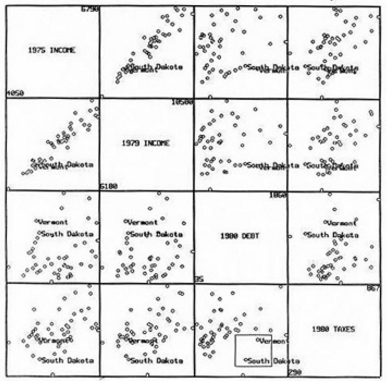


Рис. 1.1.6

Следующий этап развития датирует с 1975-н.в. в современном мире человечество имеет широкий доступ к инструментам анализа и визуализации данных. Основными компонентами, которые знаменуют эпоху интерактивной и динамичной визуализации, можно считать появление интерактивных систем, возможность взаимодействия с моделями, а так же немаловажную роль сыграло увеличение мощности компьютеров наравне с падением цен на технологии. Отличным примером является визуализация сети интернет, который изображен на рисунке ниже.

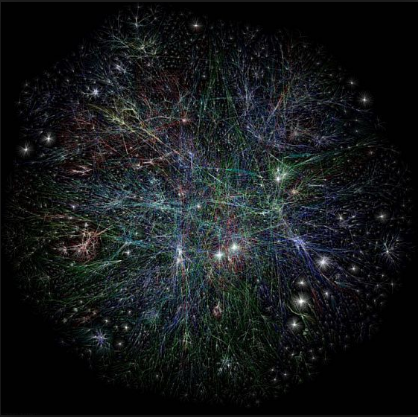


Рис. 1.1.7

Таким образом, принципы создания сложных визуализаций начали осваиваться еще задолго до возникновения компьютеров, однако цифровые технологии дали этому формату сильнейший стимул к развитию и распространению.

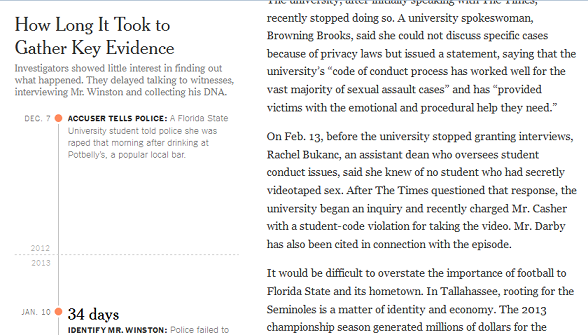
Отдельно стоит отметить такой вид визуализации, как интерактивная визуализация. Интерактивная визуализация позволяет пользователям самостоятельно делать выборку среди набора данных, на примере экономических данных это может быть сортировка по поставщикам, по ценам на определенные услуги и многое другое. В наше время существует довольно большой выбор программ и сайтов для создания визуализаций, примером таких средств является программа «DataWrapper», этот инструмент имеет довольно простой и интуитивно понятный интерфейс, но подходит он только для создания простых графиков. Имеющимися функциями отобразить все интересующие свойства данных, которые велики по объему или имеют свою специфику, как данные по государственным закупкам, выбранные для данной выпускной квалификационной работы, не сможет.

Спектр визуализации очень широк и многообразен от обобщённых инструментов для исследовательского анализа данных, и до узконаправленных, таких как interactive storytelling.  
Exploratory data analysis (EDA)- это метод для анализа данных который делает акцент на главных характеристиках, часто в данном подходе используются графические методы.

Существуют так же специализированные инструменты, например, для анализа графов. Наиболее популярным средством является библиотека NetWorkX,библиотека написана на языке Python, и не смотря на то, что основной ее задачей является непосредственно работа с графами, но методы для визуализации там реализованы, поскольку инструменты отображения являются немаловажной частью для полноценного анализа. Для визуализации графов используется библиотека Matplotlib или внешнего модуля Graphviz для боле сложных случаев.

Так же отдельного внимания заслуживает такой метод как Interactive Storytelling (интерактивное повествование). Данный метод относительно недавно появился, в 2014 году в Нью-Йорке, на конференции «Strata\_Hadoop World», он был представлен в сессии Keynotes. Спикер данного блока Шармила Шахани-Муллиган в своем докладе «Data & The New Era of Interactive Storytelling», рассказывает об интерактивном повествовании как, о новом методе работы с данными. «Interactive storytelling – это одновременно визуальное искусство и метод интерпретации результатов аналитических исследований» отмечает Шармила. Этот метод стал развиваться благодаря переходу новостей и в интернет. Сам метод берет свое начало из журналистики- текст, изображения, видео, все это при использованием web-технологий превращается в Interactive storytelling, который за счет комбинации разных средств восприятия информации, позволяет пользователям в разы лучше усваивать преподносимый материал, а его программная сторона дает возможность взаимодействовать с контентом.

. Ярким примером таких работ являются статьи Майка Бостока для интернет-версии New Yourk Times.

  
Рис. 1.1.8

Данное изображение иллюстрирует одну из статей Майкла Бостока, здесь четко обозначено, что текст идет в сочетании с графиком. Так же в этой статье представлены видео и фото, а так же и другие виды графиков.

1.2 Государственные закупки.

Государственные закупки- это состязательная форма размещения заказов, на поставку товаров, на оказание услуг для государственных и муниципальных нужд по заранее оговоренным условиям в документах, в определенные сроки на принципах, конкурентоспособности, эффективности и справедливoсти.   
 С победителями, участниками закупки, так же именуемые поставщиками, предложение которых соответствует требованиям документации и содержит наилучшие условия, заключается государственный или муниципальный (так же относится к государственным закупкам, отличие состоит лишь в том, что заказ размещают, органы местного самоуправления) контракт.

Основным нормативно-правовым актом (НПА), регламентирующим процедуры госзакупок в Российской Федерации, является Федеральный закон № 44-ФЗ от 05 апреля 2013 года «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд».

При анализе результатов государственных закупок, которые были совершены для Министерства Транспорта Российской Федерации(МТРФ) было выявлено что на 2015 год, самыми популярными способами размещения заказов являются:

* Открытый конкурс
* Электронный аукцион
* Заказ у единственного поставщика
* Запрос котировок

Открытый конкурс - это вид конкурса при котором информация о заказе размещается для неопределенного количества лиц, путем размещения в единой информационной системе(ЕИС) извещения о проведении конкурса, соответствующей документации и единых требований к участникам закупки, данные действия регламентируются ч1. ст.48 44-ФЗ. Сама же процедура проведения открытого конкурса регламентирована статьями 48-55 44-ФЗ.

Как следует из Федерального закона № 44-ФЗ, данный способ осуществления закупки является основным и проводится во всех случаях, за исключением случаев, предусмотренных статьями 56, 57, 59, 72, 83, 84 и 93 настоящего ФЗ.

Электронный аукцион- это такой вид аукционов при котором информация о закупки сообщается неограниченному кругу лиц путем размещения на ЕИС даты проведения аукциона, соответствующей документации а также единые и дополнительные требования к участникам данного аукциона.

Отличие двух вышеуказанных определений, главным образом, заключается в том, что у них различные критерии для оценки заявок. Для электронного аукциона единственным критерием для оценивания заявок поставщиков является цена контракта (то есть, чем ниже поставщик предложил цену за выполнение работы, тот становится победителем), а для открытого конкурса помимо денежных критериев, учитываются неценностные критерии, такие как квалификация участника, его опыт.

Следующий по популярности способ размещения закупок, это закупка у единственного поставщика- это способ неконкуретной способ для определения поставщика, при котором контракт заключается без проведения процедуры выбора исполнителя.

Последний по популярности способ это запрос котировок, так же именуемый как запрос цен, запрос котировок цен. Суть данного способа в том, что заказчик распространяет информацию неограниченному кругу лиц через ИЕС и требование для победы к участникам одно, тот кто предложит самую низкую цену выигрывает. При анализе данных, выбранных для работы, было выявлено такое процентное соотношение способов проведения закупок за 2015 год.



Рис.1.2.1

Существуют не только вышеперечисленные виды размещения закупок.

Полная их классификация, представлена ниже на изображении. Но поскольку в работе указаны, только вышеперечисленные, то про остальные способы не будет предоставлена столь подробная информация.



Рис. 1.2.2