**Построение карты изменения динамики преступности Украины в системе SAS ENTERPRISE GUIDE по статистическим данным МВД Украины**

Горинова Мария Александровна, студентка 3 курса

Факультета Информатики и Вычислительной Техники,

Национальный Технический Университет Украины «Киевский Политехнический Институт»

**Реферат.**

Данная научная работа предназначена для описания метода и практического применения такого вида графической визуализации изменения данных как фоновая картограмма или хороплет. В качестве входных данных были использованы статистические данные Министерства Внутренних Дел Украины, которые включают в себя общее количество преступлений и правонарушений, совершенных ежемесячно в каждом из регионов Украины в период с 2002 по 2014 год включительно. Главной задачей было построение карты Украины с последующим закрашиванием ее регионов в зависимости от принадлежности к определенному кластеру по заданным параметрам. Работа выполнялась в рамках системы SAS Enterprise Guide написанием программ на языке SAS Base и SAS SQL.

**Вступление**.

Для анализа, моделирования и статической обработки данных используют различные статистические пакеты, но наличием понятного для пользователя интерфейса могут похвастаться лишь некоторые аналитические средства. Компания SAS в инструменте SAS Enterprise Guide прежде всего объединила мощный инструмент обработки больших объемов данных и user-friendly интерфейсом по технологии drag and drop. Данный инструмент позволяет осуществлять как основные манипуляции с данными, а именно (1) доступ к базам и хранилищам данных с использованием ODBC интерфейса, (2) управление, (3) создание витрин данных с использованием SAS SQL, SAS Base и R языков программирования, (4) анализ статическими методами обработки и (5) построение аналитических отчетов, в PDF, RTF, MS Word, TXT, HTML, XML и других форматах.

**Постановка** **задачи**.

Choropleth map – фоновая картограмма, хороплет. Термин, происходящий от греческого “место, область” + “значения”. Это вид картограммы, на которой штриховкой разной частоты или красками разной степени насыщенности изображают интенсивность какого-нибудь показателя в границах территориальной единицы.

Такие карты позволяют наглядно отобразить данные по большому количеству географических объектов – областям Украины. Данный вид визуализации данных наглядно демонстрирует степень интенсивности прироста преступности в зависимости от региона.

Существует множество различных вариантов создания подобных карт. Но так как на практике редко имеется конкретная выборка данных по которой можно сразу построить хороплет, то именно поэтому SAS EG является идеальным инструментом в нашем случае, когда имеется большой массив данных, из которого нужно сделать различные выборки по определенным параметрам.

**Решение задачи.**

Пошаговое описание выполнение задачи:

Для запуска проекта необходимо на диск C скопировать содержимое папки Project\_MVD, которая включает в себя следующие данные для работы:

Набор данных Values.sas7bdat, который включает:

- 9 662 328 строк

- 6 столбцов

- 310 640 640 байт (310 МБайт)

Region - 31 уникальное значение, от 01 до 35, при этом отсутствуют 13, 14, 28 и 29.

VariableA - 16 уникальных значений, от 01 до 16.

VariableB - 126 уникальных значений, от 001 до 126.

Year - 13 уникальных значений, от 2002 до 2014 года.

Month - 12 уникальных значений, от 1 до 12.

Amount - 7 573 уникальных значений

Набор данных Location

- 33 строки

- 3 столбца

Code – код области - 33 уникальных значения, от 1 до 35, отсутствуют 28 и 29.

Region – регион - название региона.

Всего возможно 60 016 комбинаций (временных рядов).

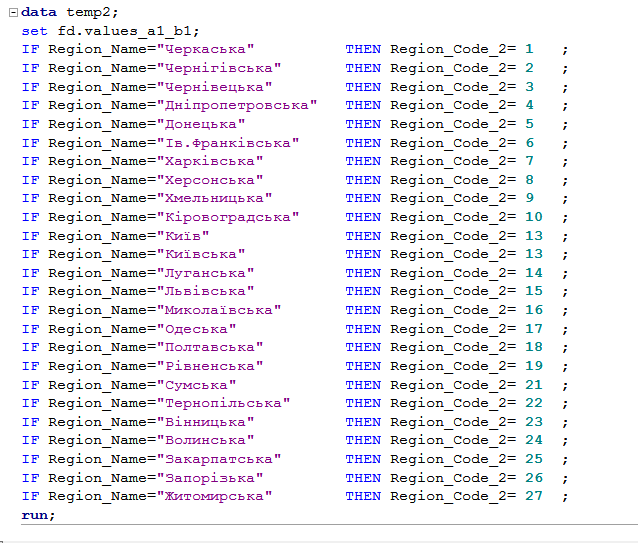
Перед началом работы для удобства работы с небольшими объемами данных было выбрано среди всех записей, что есть в начальной выборке данных МВД только те, которые относятся к регионам по таким параметрам:

Region\_Code le 27 – меньше 27, так как имеется 31 уникальное значение, от 01 до 35, при этом отсутствуют 13, 14, 28 и 29, VariableA\_Code=1 – количество зарегистрированных преступлений и криминальных правонарушений, VariableB\_Code=1 – все типы преступлений и криминальных правонарушений. Выбираем

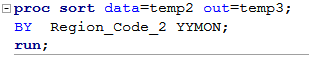
*Первый этап.* Подготовка выборки данных.



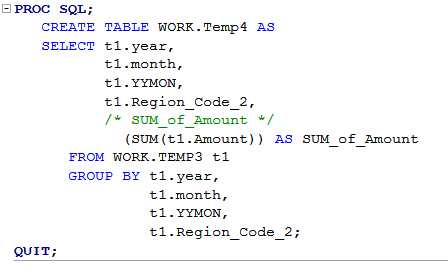
Подключение библиотек fd и mymap, в которых хранятся начальный набор данных и данные для построения карты соответственно.



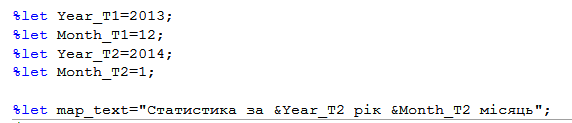
На этом шаге данных выполняется перекодирование регионов с созданием новой переменной Region\_Code\_2.



Процедурный шаг, в котором данные сортируются по региону и году-месяцу для того, чтобы в следующем коде программы выполнить суммирование количества преступлений, совершенных в один месяц для города и его области, потому что на карте город не отображается, поэтому совершенные преступления должны быть отнесены к соответствующей области.

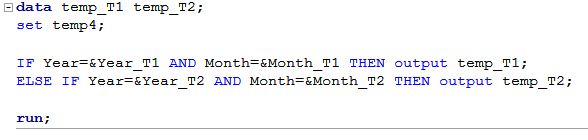


Процедурный шаг SQL создает таблицу, в которую включаются колонки и по ним же группируются: год, месяц, месяц и год в формате YYMON и код региона, а также создается новая колонка, в которой рассчитывается показатель прироста числа совершенных преступлений суммируя количество всех преступлений по одному региону.



Макропеременные Year\_T1, Month\_T1, Year\_T2 и Month\_T2 используются для определения предыдущего и текущего месяцев и соответствующих годов по которым рассчитывается прирост числа совершенных преступлений.

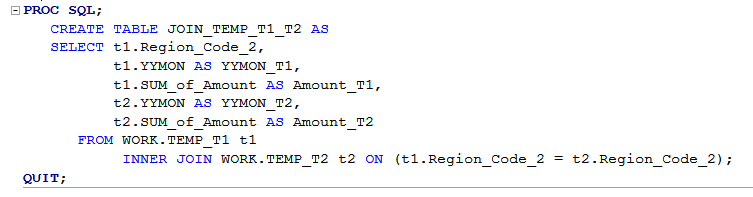
Макропеременная map\_text используется для создания подписи под картой Украины.



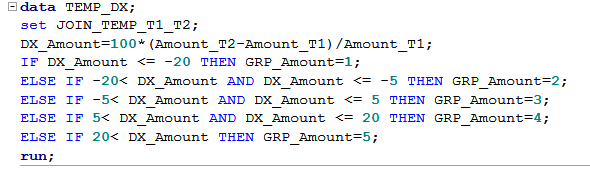
Эта часть кода программы предназначена для формирования наборов данных temp\_T1 и temp\_T2.

temp\_T1 - набор данных, который содержит количество преступлений, совершенных в предыдущий месяц в разрезе регионов Украины, temp\_T2 содержит количество преступлений в текущий месяц.

На основе этих наборов данных будет рассчитано изменение (прирост или уменьшение) числа преступлений для соответствующего региона.



SQL запрос для объединения статистики о числе совершенных преступлений в предыдущий и текущий месяц для каждого региона.



Код программы для расчета прироста числа преступлений для соответствующего региона.

DX\_Amount - разница между количеством преступлений в предыдущем и текущем месяце.

GRP\_Amount - категориальная переменная числового типа содержащая значения:

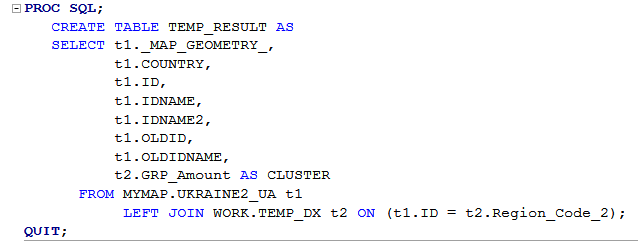
1 - количество преступлений упало больше чем и на 20%;

2 – количество преступлений упало в диапазоне от 5% включительно до 20%;

3 – количество преступлений осталось почти неизменным, упало в диапазоне от 0% до 5 %, либо возросло от 0% до 5% включительно;

4 - количество преступлений возросло в диапазоне от 5% включительно до 20% включительно;

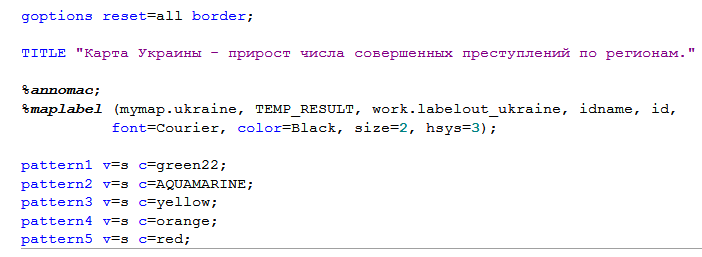
5 - количество преступлений возросло больше чем и на 20%.



Код программы для создания таблицы данных, на основе которой будет выполнятся построение карты Украины с заливкой регионов цветом в зависимости от принадлежности к кластеру (GRP\_Amount).

*Второй этап*. Построение карты.

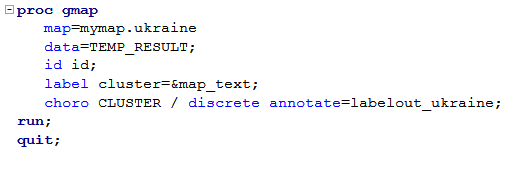
Теперь на основе информации из TEMP\_RESULT нужно выполнить закраску регионов.



Код программы, который выводит карту и заголовок.

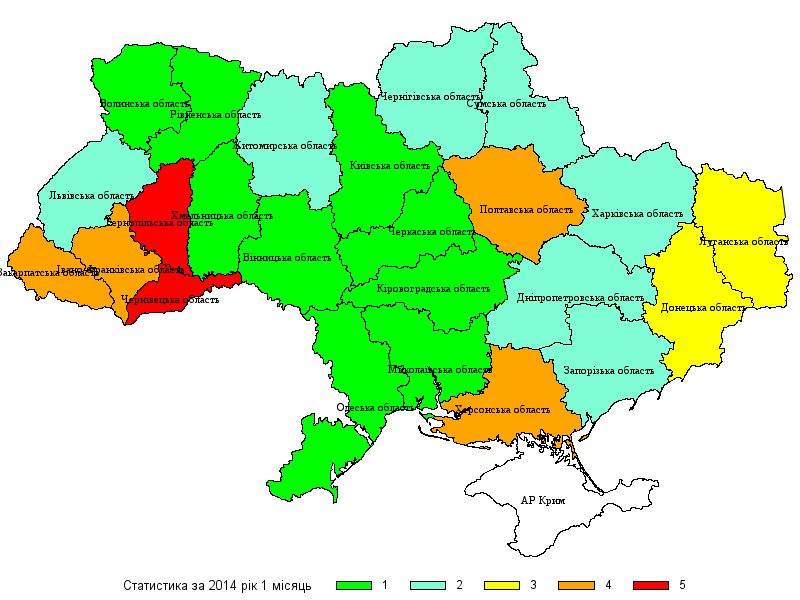
%annomac – это макрос, который запускают, перед тем как использовать аннотации в макросах, которые используются в дальнейшем, %maplabel – макрос, предназначенный для создания файла с текстом который будет отображаться на карте для каждого региона, а также определения стиля его вывода.

Оператор pattern используется для определения цвета заливки регионов в зависимости от принадлежности кластеру.



Выполняется процедура gmap для построения карты Украины, в котором указываются роли переменных при построении карты.

Результат:



[Ссылка](https://1bd09614-a-62cb3a1a-s-sites.googlegroups.com/site/sas4ua2015/gorinova_maria/MAPS_UKRAINE_MVD.gif?attachauth=ANoY7crFLsmsma4zxsfyDByuahirowoULJSCMXSHGlmZFTI7FvPkK1pxgKe7q80RYoyiUKXtC3vfuxhhMB8R9d-H0Z1EkC_PL7yovkgwyHqQ1G5cTOTsalNpBxzI5Sb3zAcZW6pblx3FMWTnUucVtcJnxhieAtO6IwQMIRx_oheReAnZVtIDQjk7t7ci2BvYkCjgThrw3wxbGMTWRyUhFL876i00-GQiU25HOt-UyX1hmbV2iVm5tC8%3D&attredirects=0), по которому можно просмотреть GIF-анимированное изображение, которое включает все результаты (с 2012 по 2014 год).

**Выводы.**

Фоновая картограмма или хороплет является самым наглядным и удобным решением для графической визуализации интенсивности определенного параметра в переделах выбранных географических единиц, особенно если их много. В результате выполненного в данной работе построения карты по статистическим данным преступности, когда временные промежутки прироста сравнительно небольшие, а регионов такое количество, которое не располагает к детальному изучению динамики изменения по каждой области, использование хороплета с привычным раскрашиванием по интенсивности в спектре от зеленого (благоприятная ситуация) до красного (критическая ситуация) с промежуточными цветами, не только позволяет легко и быстро оценить динамику прироста преступности для каждого из регионов, но и сделать сравнение карт с результатами по другим промежуткам времени для виденья картины в целом, например за год или больше.

**Источники.**

SAS Documentation. Enterprise Guide - <http://support.sas.com/documentation/onlinedoc/guide/>;

Терентьев А.Н., Домрачев В.М., Костецкий Р.И. SAS Base: Основы программирования. - К. : Эдельвейс, 2014 – 304 с.

**Контактная информация.**

Горинова Мария Александровна, email: [mariagorinovaa@gmail.com](mailto:mariagorinovaa@gmail.com).

Факультет Информатики и Вычислительной Техники,

Национальный Технический Университет Украины «Киевский Политехнический Институт», Украина, 03056, г. Киев, проспект Победы 37.