

# **A Look Into New York**

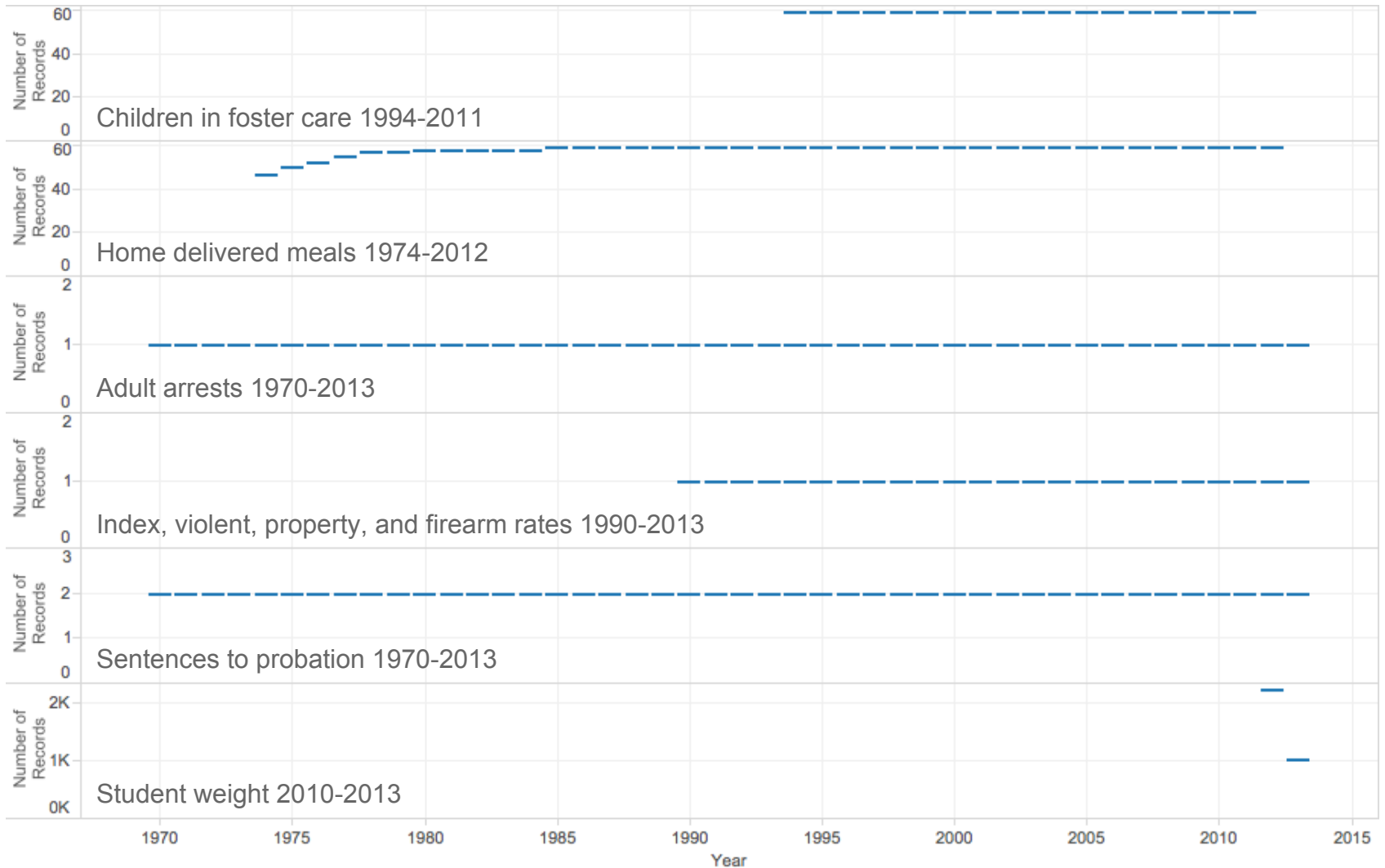
Crime, Foster Care, and Food Delivery Statistics

Bruno De Hoyos  
Estevan Garza

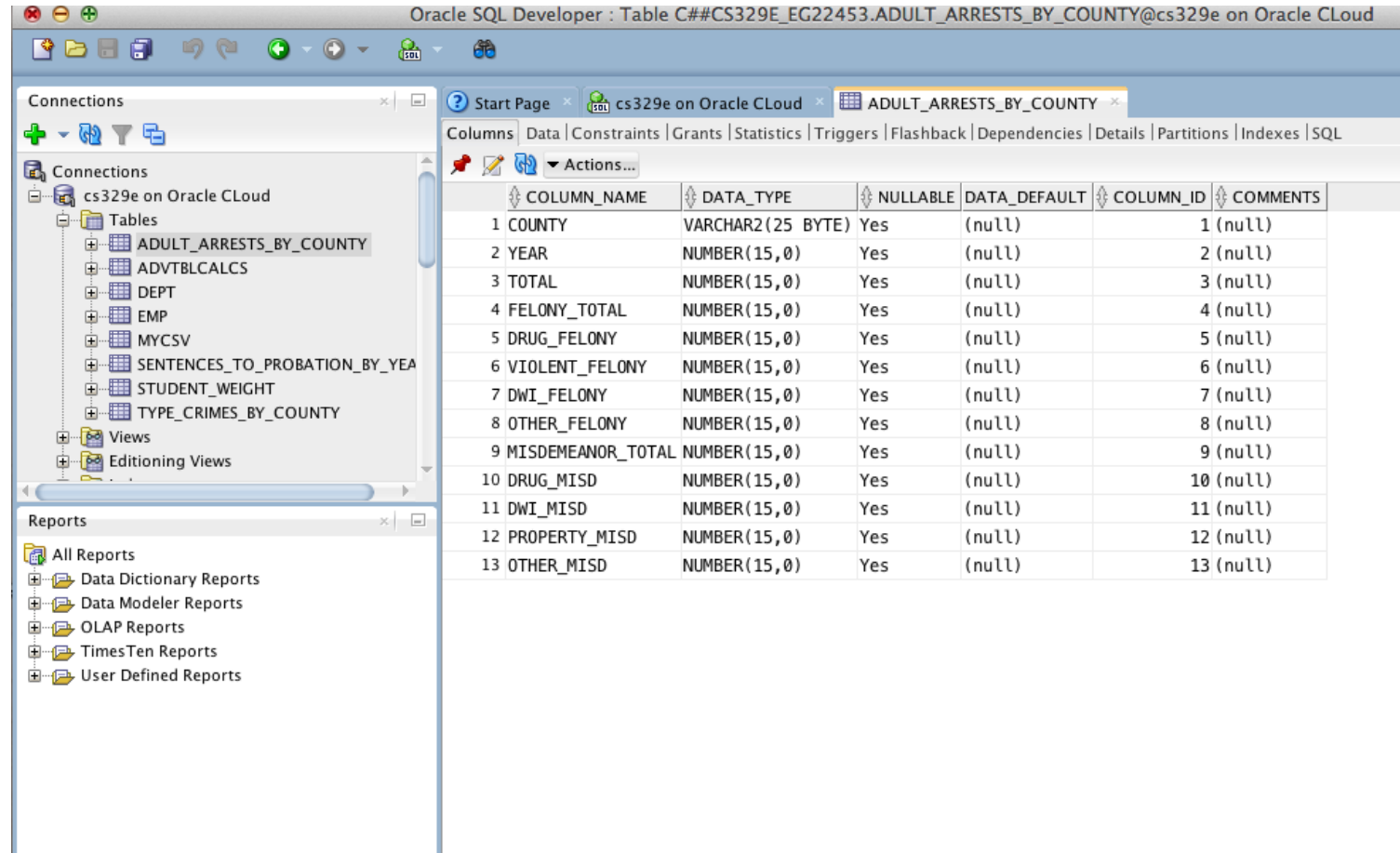
# Datasets

- Children in foster care 1994-2011
- Home delivered meals 1974-2012
- Adult arrests 1970-2013
- Index, violent, property, and firearm rates 1990-2013
- Sentences to probation 1970-2013
- Student weight 2010-2013

# Datasets - Timeline



# Loading CSV into Oracle SQL Developer



The screenshot shows the Oracle SQL Developer interface. The title bar indicates the connection is 'cs329e on Oracle CCloud'. The left pane shows the 'Connections' tree with 'cs329e on Oracle CCloud' selected, and the 'Tables' folder expanded, showing 'ADULT\_ARRESTS\_BY\_COUNTY'. The right pane shows the 'Columns' tab for this table, displaying a list of 13 columns with their data types, nullability, and default values.

	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1	COUNTY	VARCHAR2(25 BYTE)	Yes	(null)	1 (null)	
2	YEAR	NUMBER(15,0)	Yes	(null)	2 (null)	
3	TOTAL	NUMBER(15,0)	Yes	(null)	3 (null)	
4	FELONY_TOTAL	NUMBER(15,0)	Yes	(null)	4 (null)	
5	DRUG_FELONY	NUMBER(15,0)	Yes	(null)	5 (null)	
6	VIOLENT_FELONY	NUMBER(15,0)	Yes	(null)	6 (null)	
7	DWI_FELONY	NUMBER(15,0)	Yes	(null)	7 (null)	
8	OTHER_FELONY	NUMBER(15,0)	Yes	(null)	8 (null)	
9	MISDEMEANOR_TOTAL	NUMBER(15,0)	Yes	(null)	9 (null)	
10	DRUG_MISD	NUMBER(15,0)	Yes	(null)	10 (null)	
11	DWI_MISD	NUMBER(15,0)	Yes	(null)	11 (null)	
12	PROPERTY_MISD	NUMBER(15,0)	Yes	(null)	12 (null)	
13	OTHER_MISD	NUMBER(15,0)	Yes	(null)	13 (null)	

Manual data cleaning + Cleaning in RStudio

# Query into RStudio

```
# import the 'Crime Types by County' data set
dft <- data.frame(fromJSON(getURL(URLEncode('129.152.144.84:
5001/rest/native/?query="select * from
TYPE_CRIMES_BY_COUNTY"'),httpheader=c(DB='jdbc:oracle:thin:
@129.152.144.84:1521:ORCL', USER='C##cs329e_eg22453',
PASS='orcl_eg22453', MODE='native_mode', MODEL='model',
returnDimensions = 'False', returnFor = 'JSON'), verbose =
TRUE)))
```

# Joining Tables in RStudio

```
# Join two of the tables together into a "temp" table
temp <- full_join(dfs2, dfr2, by = c("COUNTY", "YEAR"))

# Now join the "temp" table to the last crime data set
full <- full_join(temp, dft, by = c("COUNTY", "YEAR"))

# Next we use complete.cases function to remove any NA's
cleantbl <- full[complete.cases(full), ]
```

# Wrangling

```
# Filter out the top 50th percentile of the data with  
respect to DWI misdemeanors.
```

```
wr1 <- cleantbl %>% mutate(misd_pct = cume_dist(DWI_MISD)) %  
>% filter(misd_pct <= 0.5)
```

```
# Filter out schools in the top 10th percentile with respect  
to overweight or obese children.
```

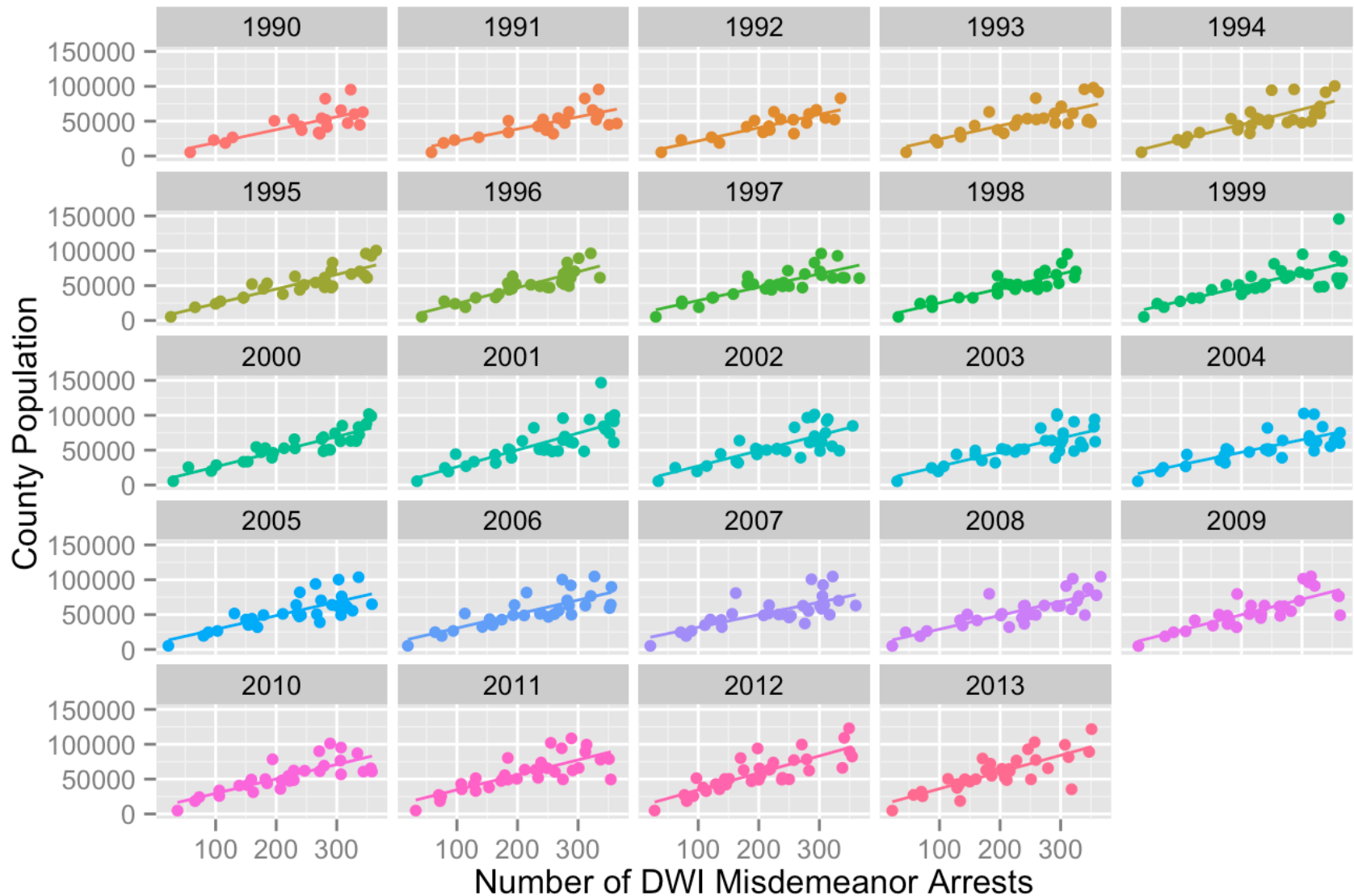
```
st1 <- student_weight %>% mutate(ow_ob_pct =cume_dist  
(NO_OVERWEIGHT_OR_OBESE)) %>% filter(ow_ob_pct <= 0.9,  
GRADE_LEVEL != "DISTRICT TOTAL")
```

# Visualizations

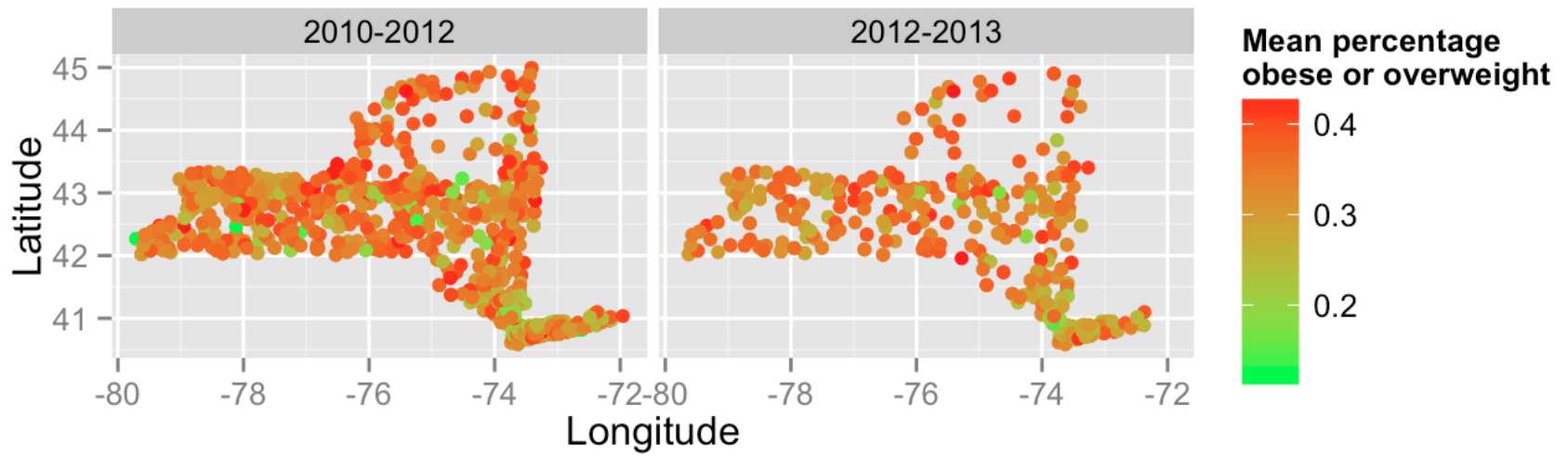
with RStudio



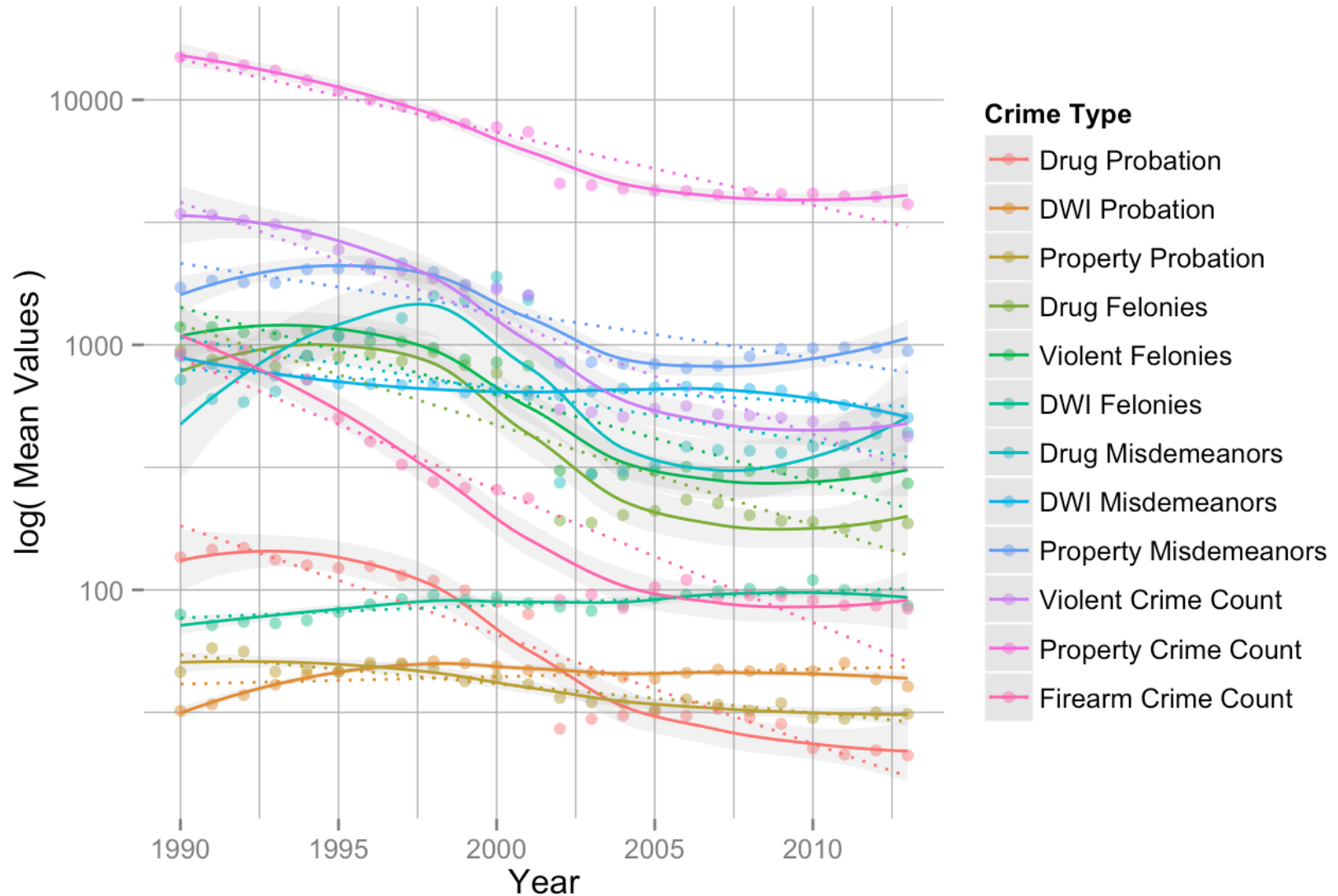
# DWI Misdemeanor Arrests vs NY County Population by Year



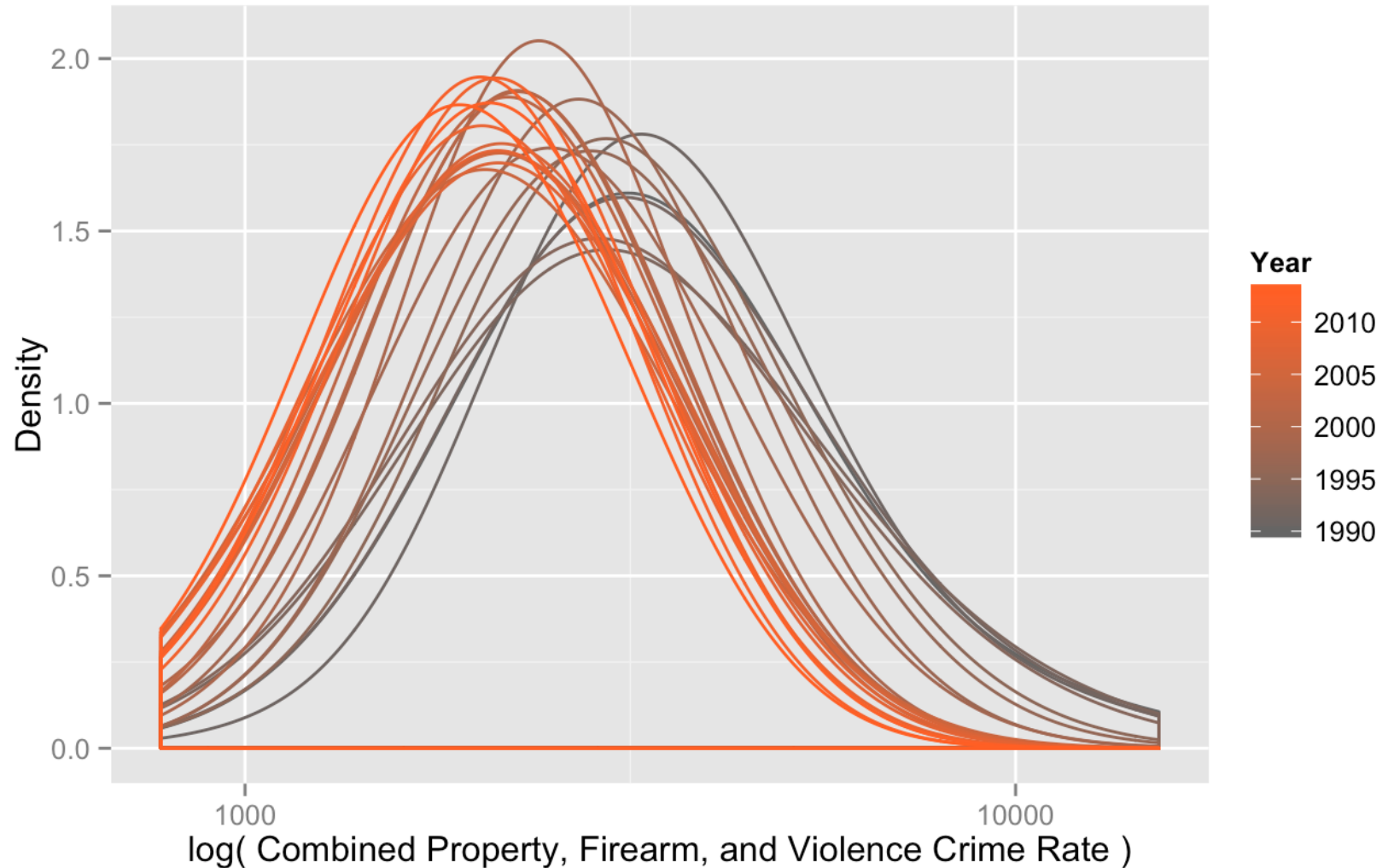




# New York Crime Averages Per Year



# New York Property, Firearm, and Violence Crime Rate Density Distributions Per Year



# **Visualizations**

with Tableau

# Importing Data into Tableau

**Connect**

**In a file**

- Tableau Data Extract
- Microsoft Excel
- Text File
- Import from Workbook

**On a server**

- Tableau Server
- Amazon EMR
- Amazon Redshift
- Cloudera Hadoop
- Firebird
- Google Analytics
- Google BigQuery
- Hortonworks Hadoop Hive
- HP Vertica
- Microsoft SQL Server
- MySQL
- OData
- Oracle
- Pivotal Greenplum Database
- PostgreSQL
- Salesforce
- Spark SQL
- Teradata
- Windows Azure Marketplace

**Oracle**

Server: 129.152.144.84

Service: orcl.usuniversity01134.oraclecloud.internal Port: 1521

Enter information to sign in to the server:

☐ Use Windows Authentication

☒ Use a specific username and password:

Username: C##cs329e\_eg22453

Password: .....

**Connect**

Server: 129.152.144.84

Service: ORCL.usuniversity01134.oraclecloud.internal

Port: 1521

User: C##cs329e\_eg22453

Pass: orcl\_eg22453

# Data Blending Relationships

Relationships determine how data from secondary data sources are joined with primary data sources.

Primary data source:  
SENTENCES\_TO\_PROBATION\_BY\_YEAR (C##CS329E\_EG22453)

Secondary data source:  
ADULT\_ARRESTS\_BY\_COUNTY

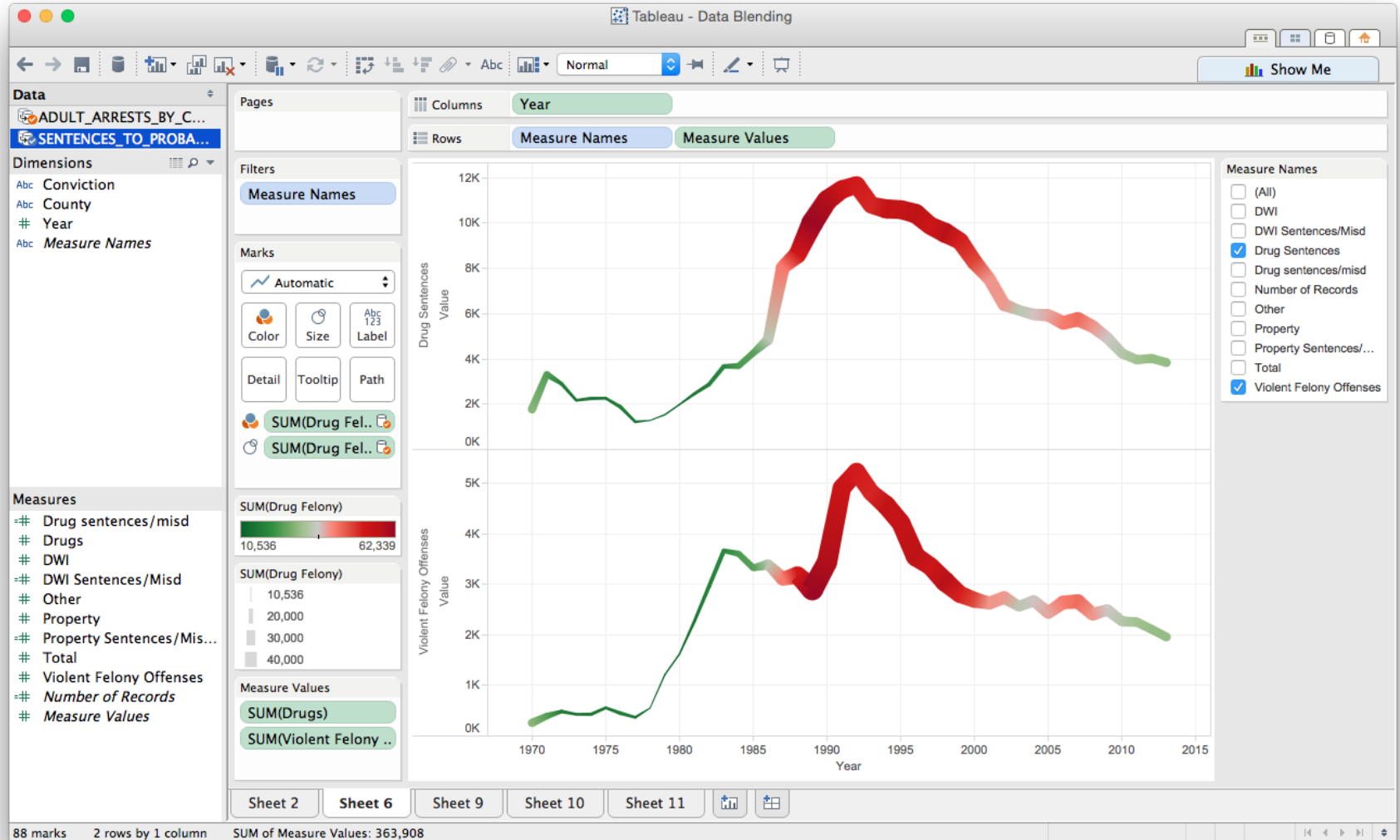
☒ Automatic ☐ Custom

County	County
Year	Year

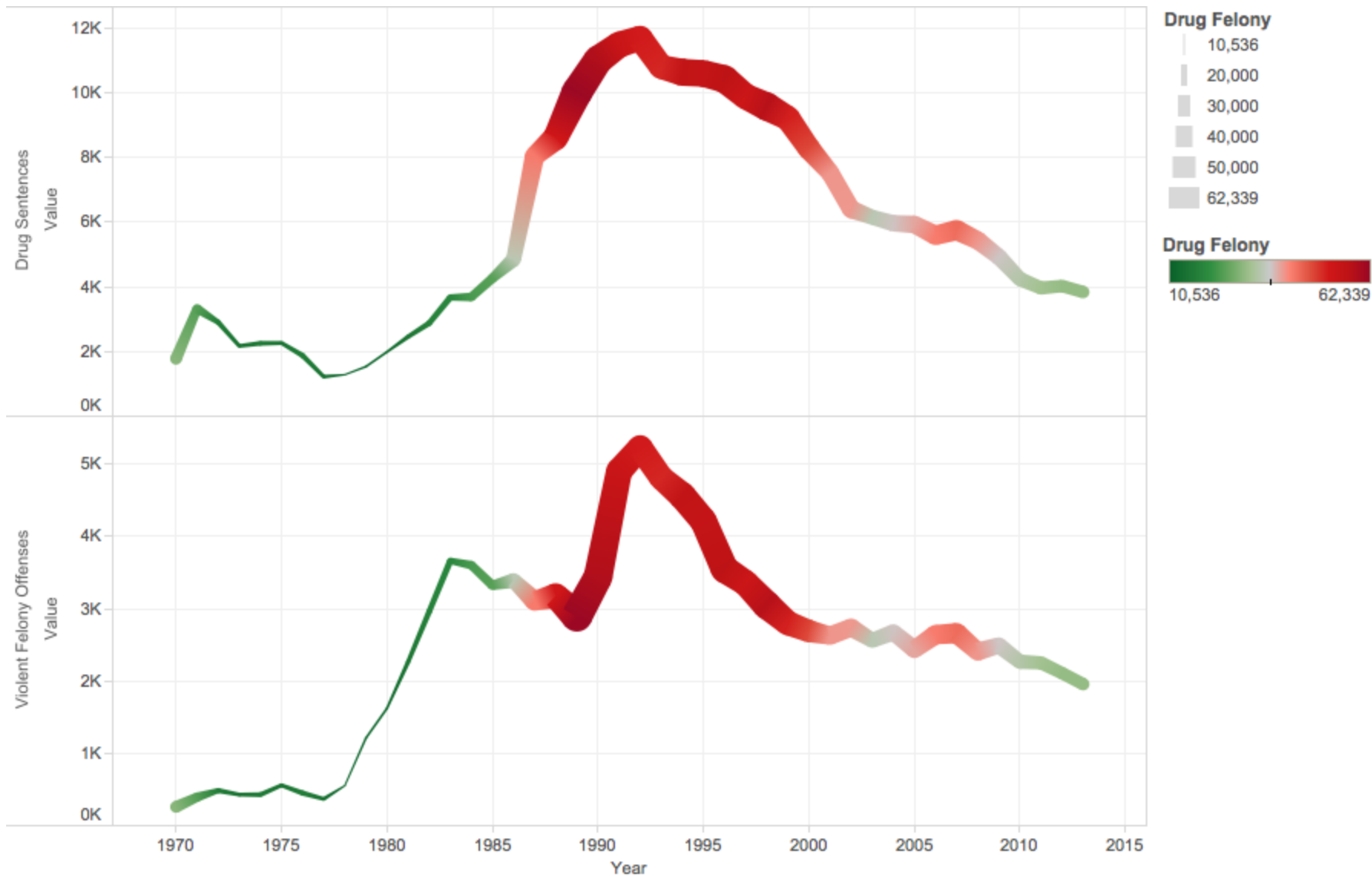
Add... Edit... Remove

Cancel OK





2 Data sources: Adult arrests 1970-2013; Sentences to probation 1970-2013



2 Data sources: Adult arrests 1970-2013; Sentences to probation 1970-2013

# Calculated Fields

Calculated Field [Drug sentences/misd]

Name: Drug sentences/misd

Formula:  
`Sum([Drugs])/SUM([ADULT_ARRESTS_BY_COUNTY (C##CS329E_EG22453)].[Drug Misd])`

✓ The calculation is valid. Clear

Fields:  

SENTENCES\_TO\_PROBATION\_BY\_YEAR (C##CS329E\_EG22453)

Enter Text to Search  
Conviction  
County  
Drugs  
DWI  
DWI Sentences/Misd  
Number of Records  
Other  
Property  
Property Sentences/Misd Arrests  
Total  
Violent Felony Offenses  
Year

Parameters: Create  

All

Enter Text to Search

Functions: Help  

All

Enter Text to Search  
SCRIPT\_BOOL  
SCRIPT\_INT  
SCRIPT\_REAL  
SCRIPT\_STR  
SIGN  
SIN  
SIZE  
SPACE  
SQRT  
SQUARE  
STARTSWITH  
STDEV  
STDEVP  
STR  
SUM

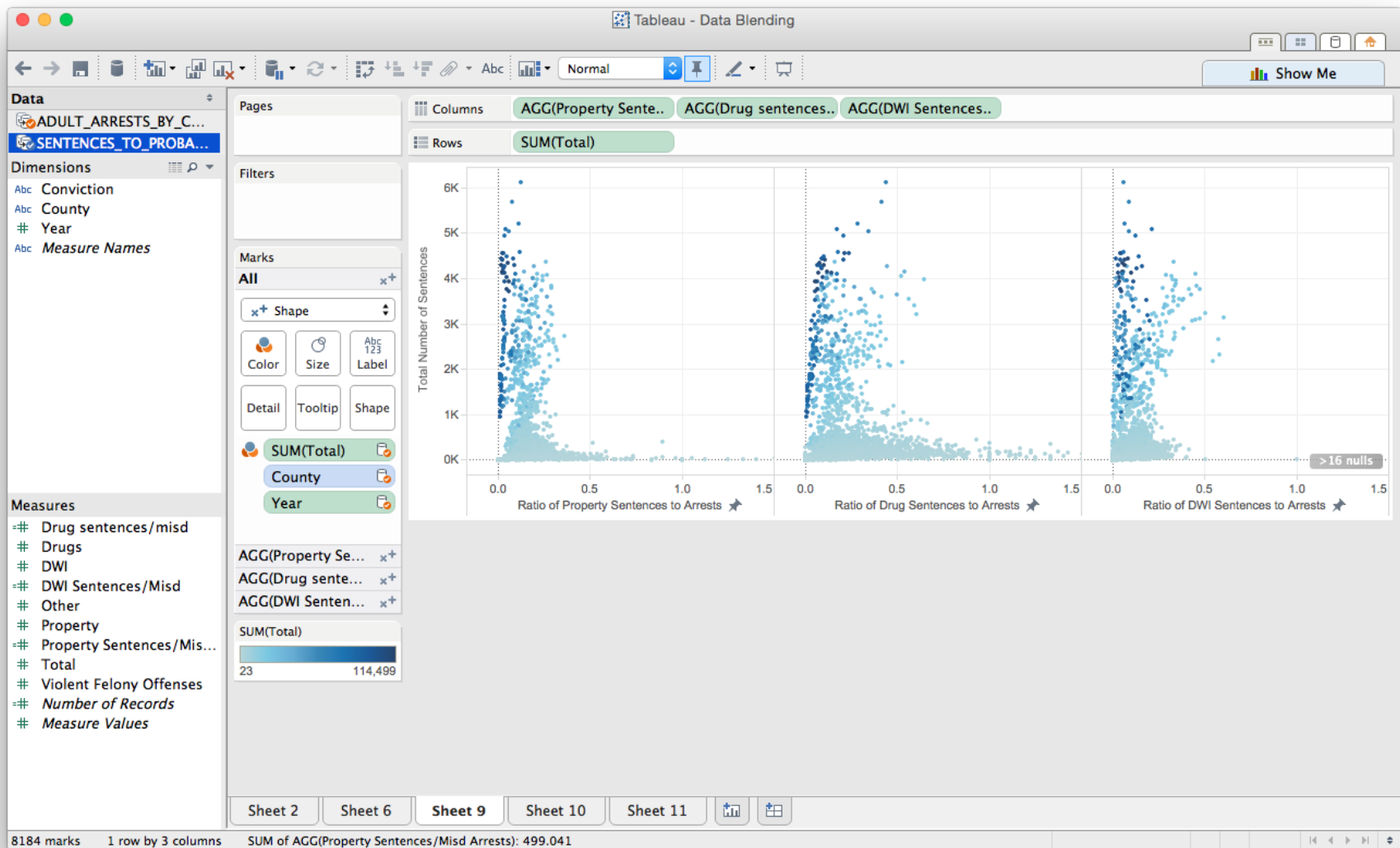
**SUM(expression)**  
Returns the sum of all the values in the expression. SUM can be used with numeric fields only. Null values are ignored.

Changes to this calculation might change the following:  
Worksheet "Sheet 9"

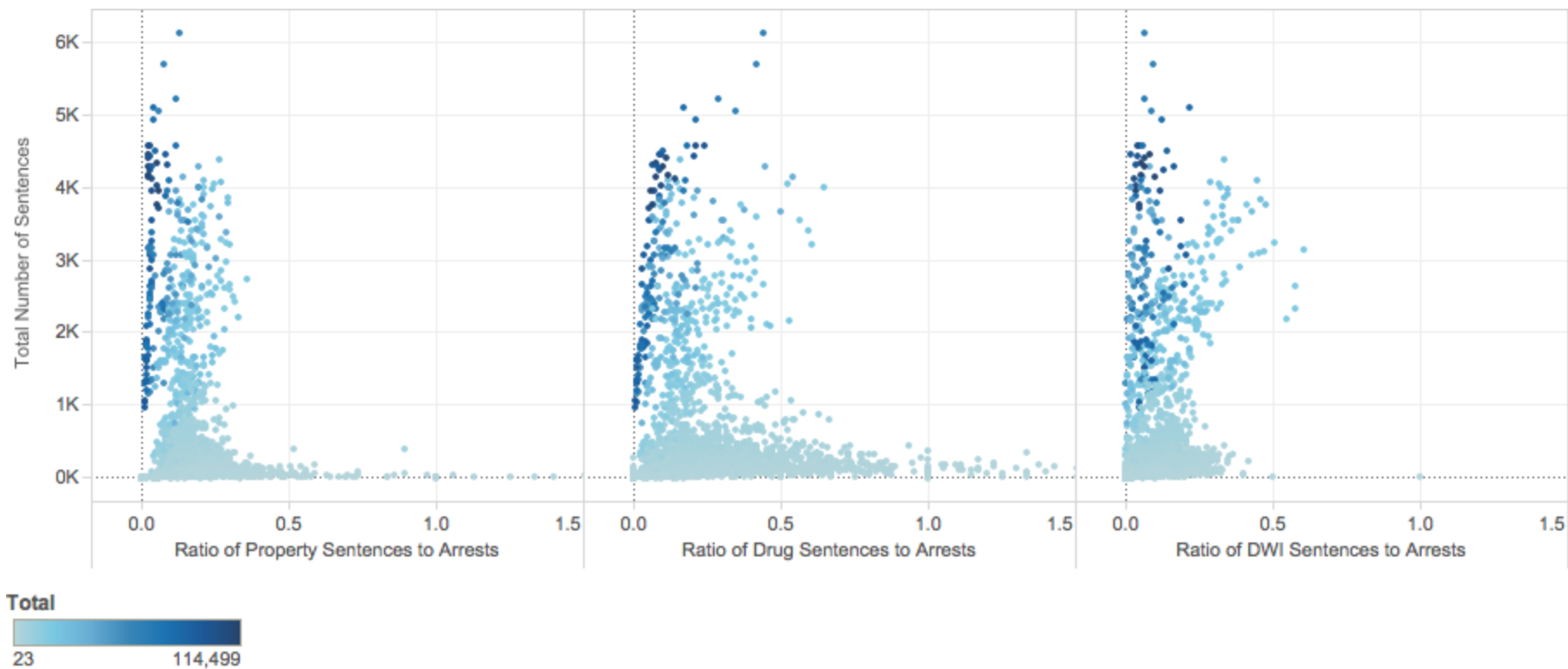
Apply

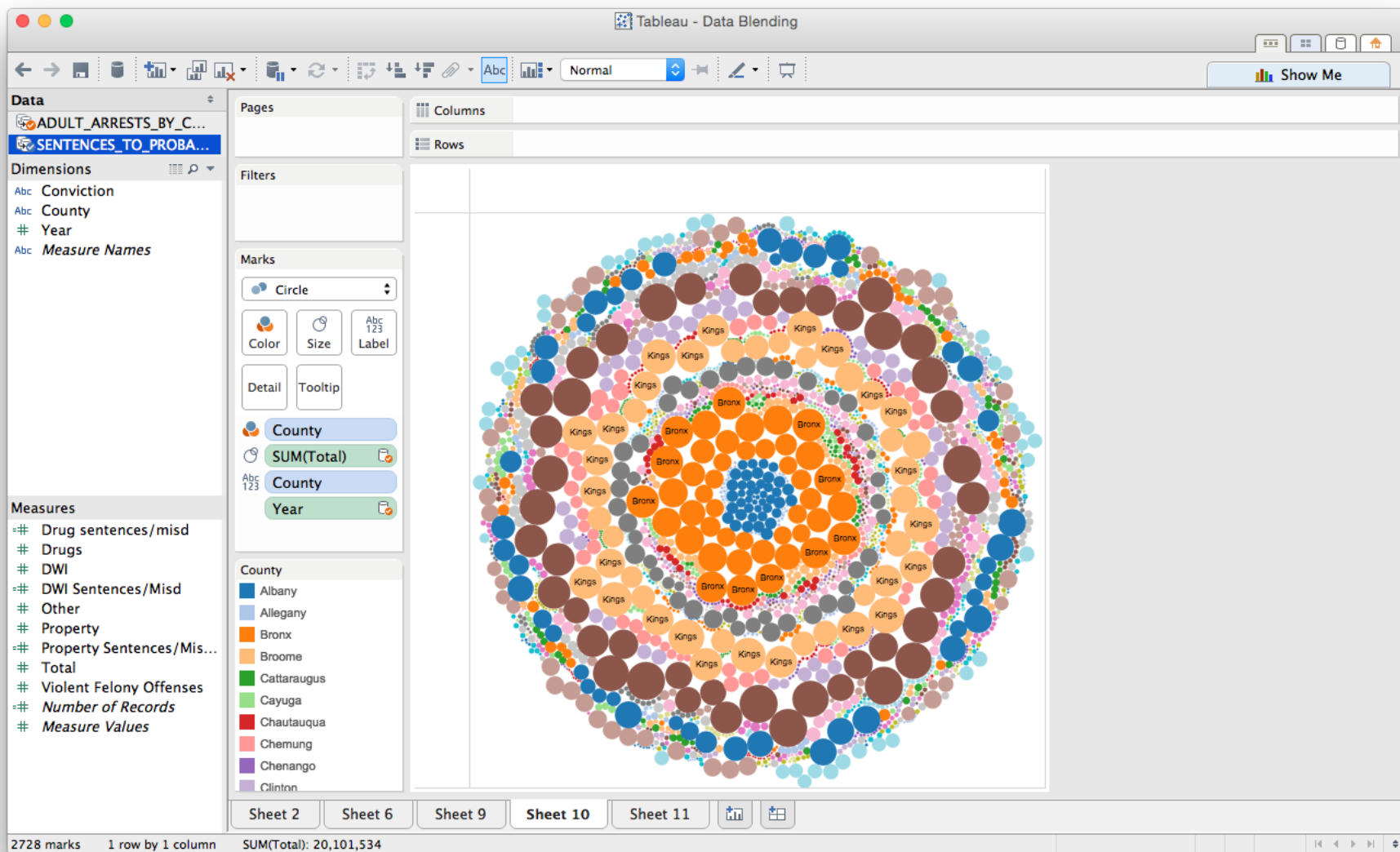
Cancel

OK



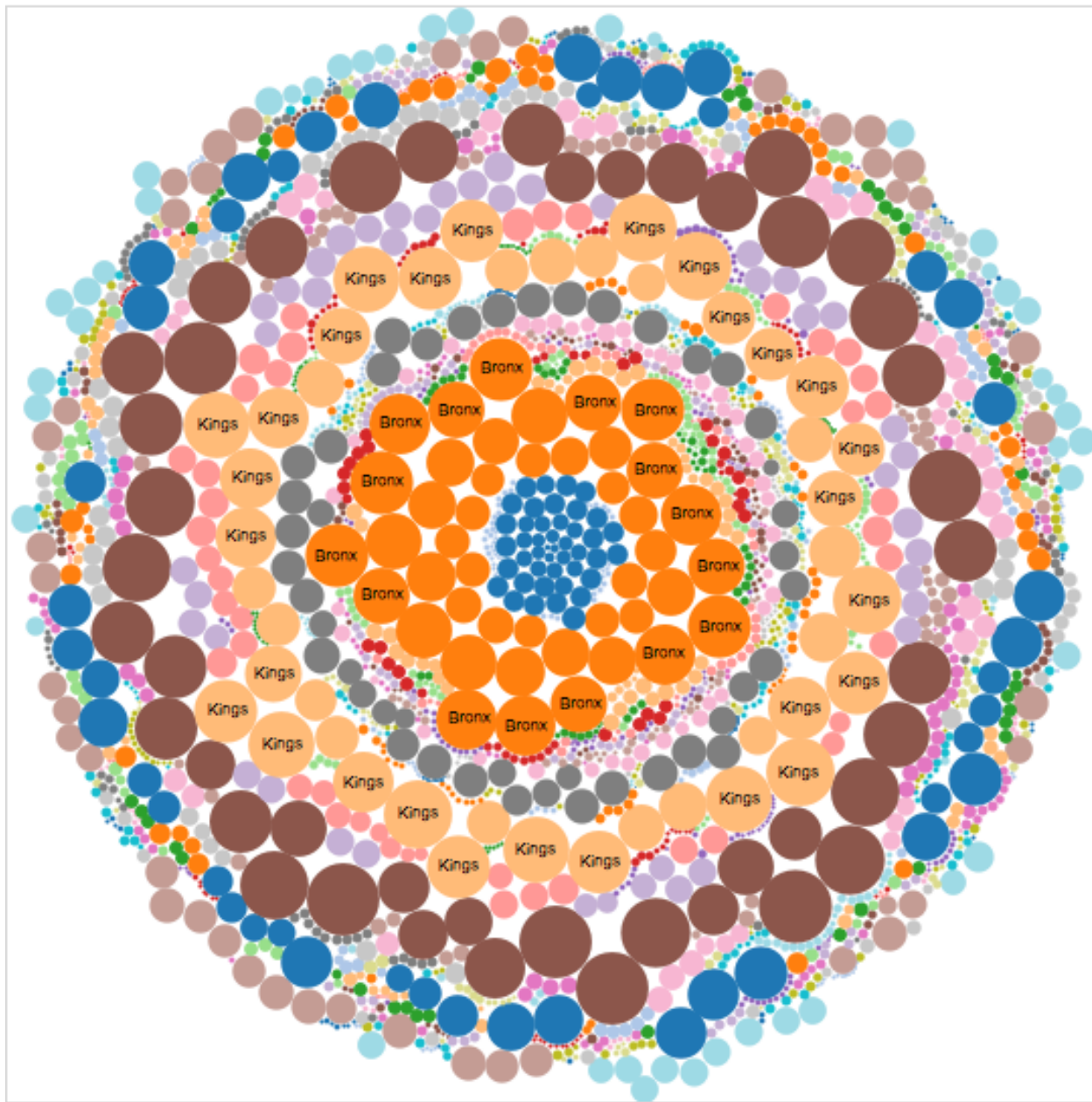
2 Data sources: Adult arrests 1970-2013; Sentences to probation 1970-2013



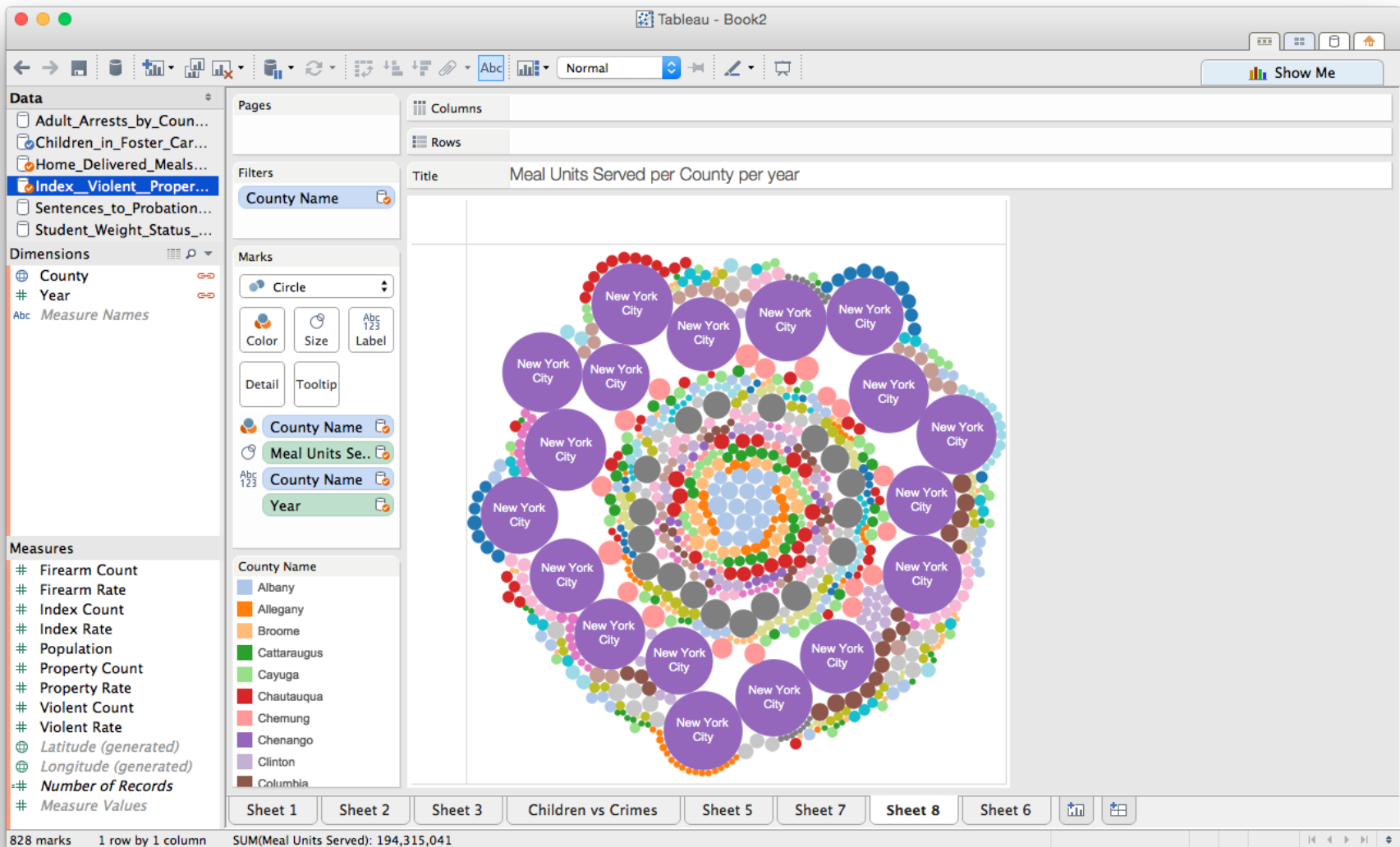


2 Data sources: Adult arrests 1970-2013; Sentences to probation 1970-2013

## Total Arrests per County per Year

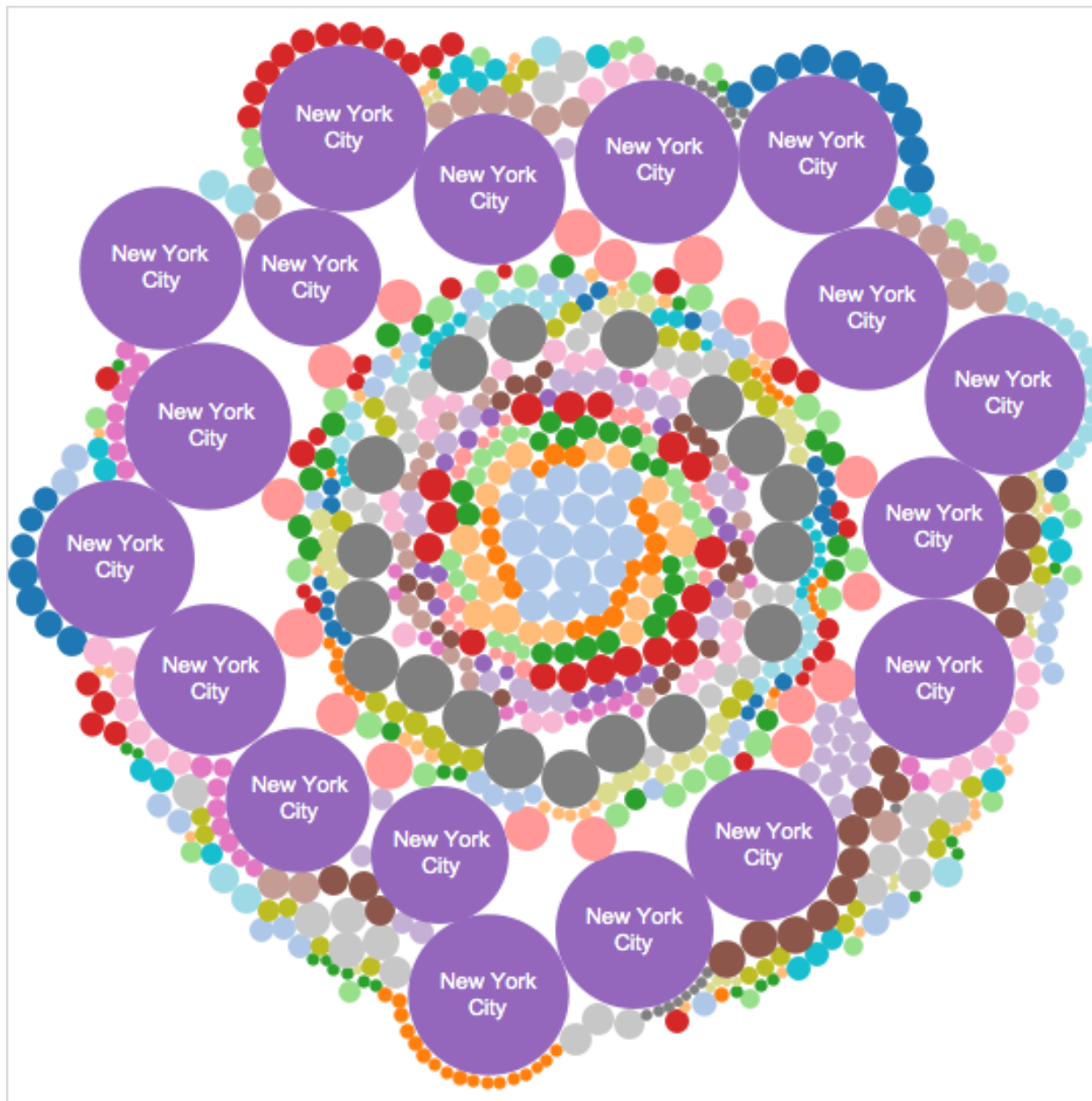


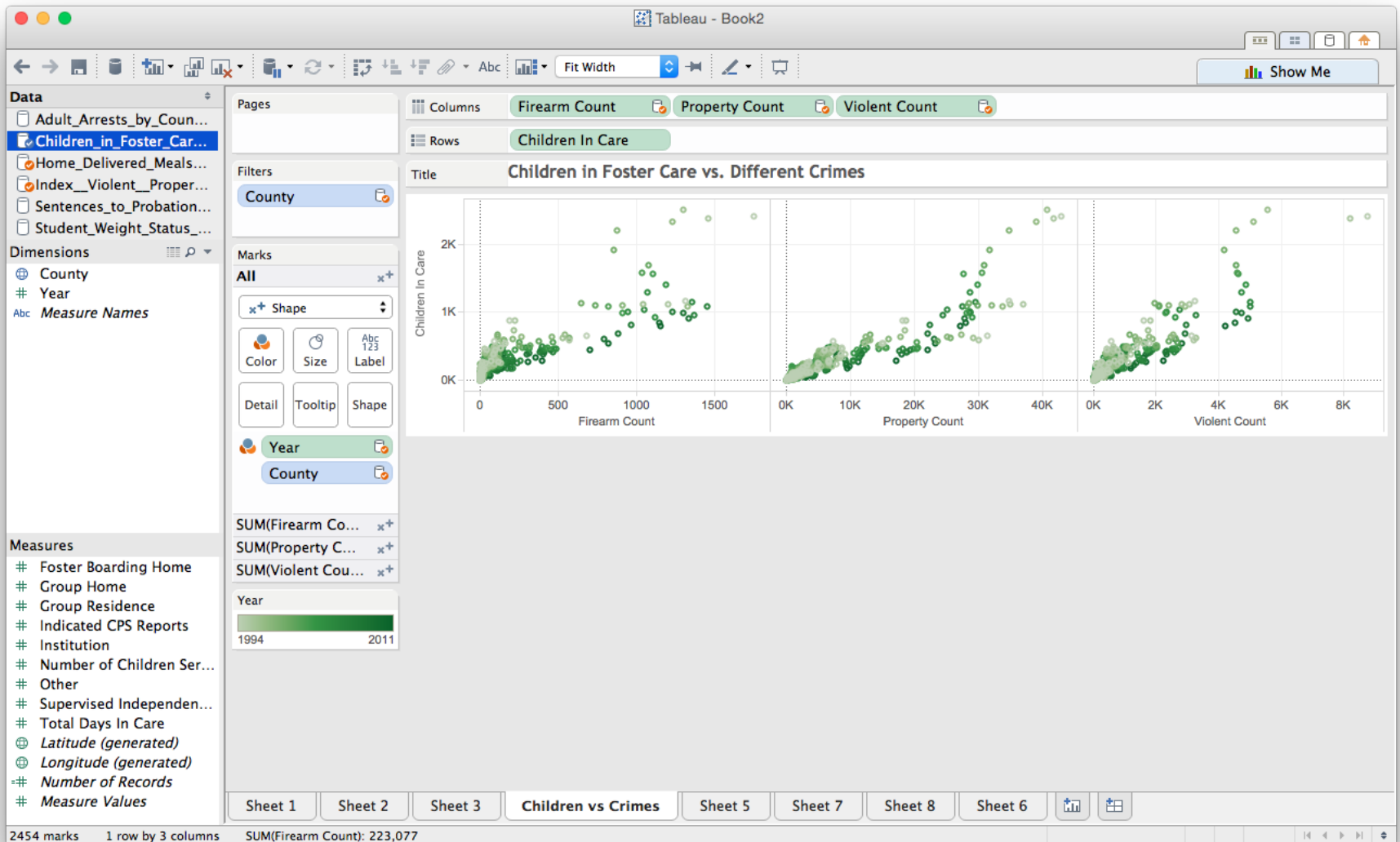
2 Data sources: Adult arrests 1970-2013; Sentences to probation 1970-2013





## Meal Units Served per County per year





2 Data sources: Children in foster care 1994-2011; Index, violent, property, and firearm rates 1990-2013

## Children in Foster Care vs. Different Crimes



2 Data sources: Children in foster care 1994-2011; Index, violent, property, and firearm rates 1990-2013

# Observations

- Expected relationships:
  - Obese vs overweight students
  - DWI arrests vs population
  - Crime rates per year
  - Drug sentences vs drug felonies vs violent felonies
  - Ratio of crime sentences to arrests
  - Children in foster care vs crime rates
- Unexpected results:
  - percentages of obesity across NY state “randomly” distributed

# Github URL

[https://github.com/eg22453/DV\\_FinalProject.git](https://github.com/eg22453/DV_FinalProject.git)

**Questions?**