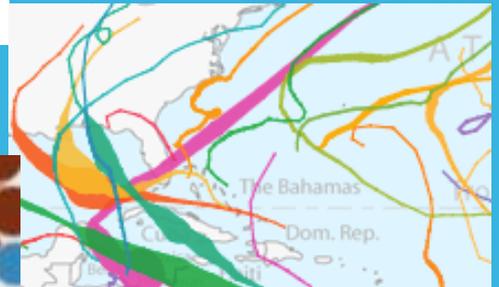
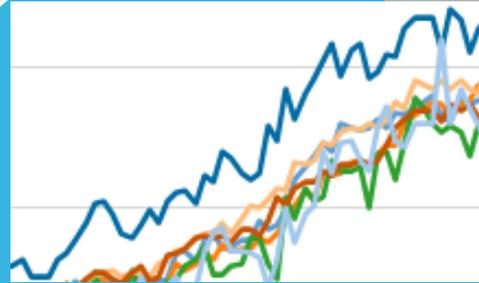


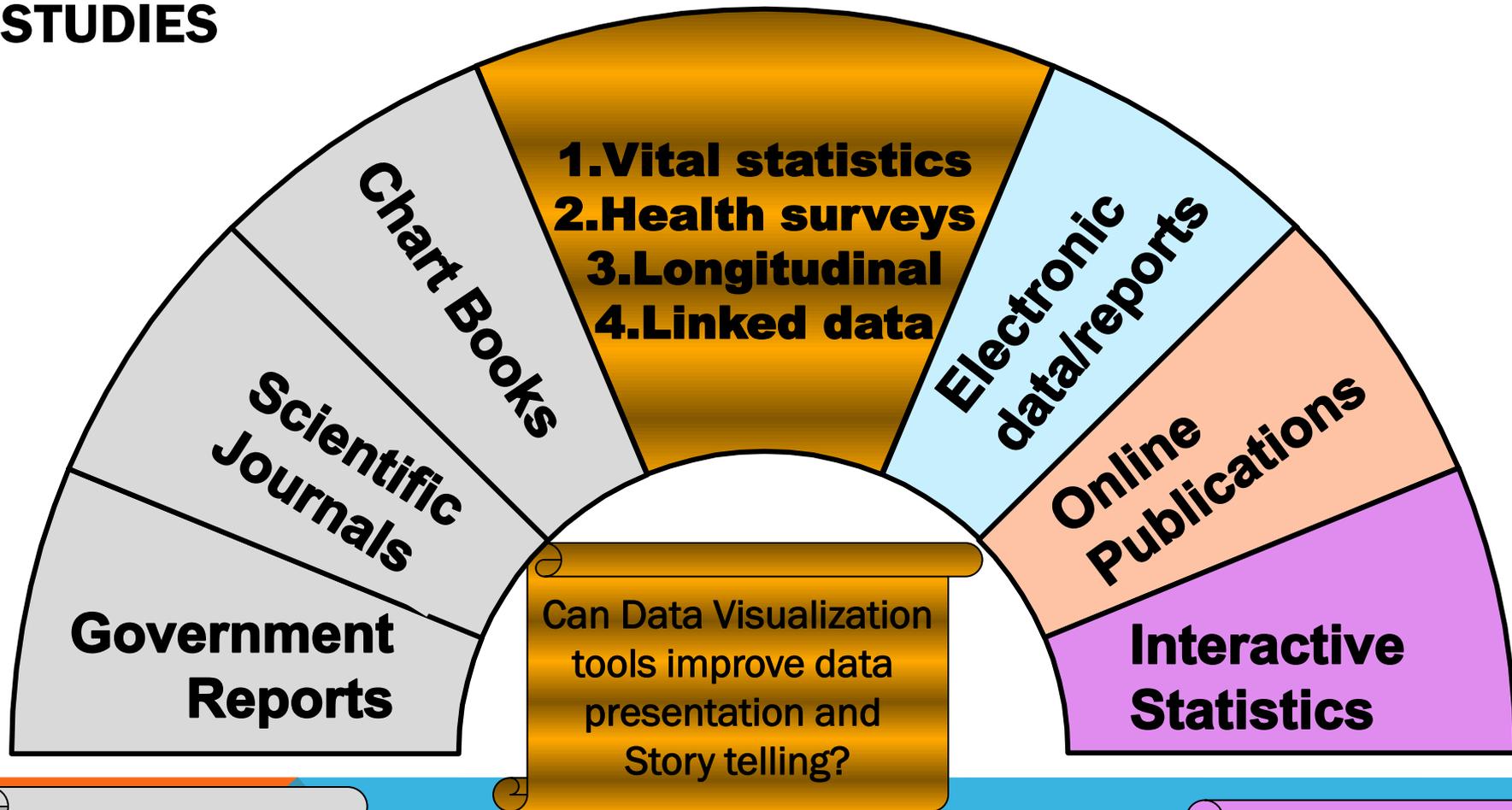
# Visual Analytics for Discovery and Insight

NCHS EXPERIENCE - YINONG CHONG, PHD



July-Nov, 2014

# NCHS DATA & PUBLICATIONS: GOLD MINE FOR HEALTH STUDIES



## Print version:

- Government reports
- Peer reviewed journals
- Chart books

## Electronic files:

- Data downloads
- PDF files
- Excel tables

## Online publication:

- Health E Stats
- Data Briefs
- Quick Stats

## Interactive stats/ query tables:

- Beyond 2020
- Health Indicators
- DOCQ

# THE WAY WE PRESENT DATA: EVOLVING WITH TIME

## Advance Data

From Vital and Health Statistics

Number 395 • May 20, 2008



### Healthy Eating Index Scores Among Adults, 60 Years of Age and Over, by Sociodemographic and Health Characteristics: United States, 1999–2002

by R. Bethene Ervin, Ph.D., R.D., Division of Health and Nutrition Examination Surveys

#### Abstract

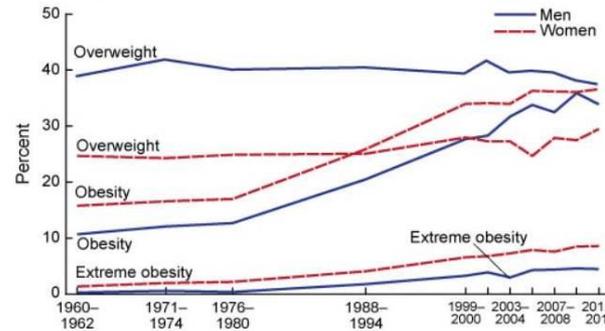
**Objective**—This report presents Healthy Eating Index (HEI) scores for adults, 60 years of age and over, from the National Health and Nutrition Examination Survey (NHANES), 1999–2002, and examines the association between the HEI scores and sex, race, and education.

#### Introduction

Diet and nutrition play important roles in maintaining health and preventing disease (1,2). This is



Figure. Trends in adult overweight, obesity, and extreme obesity among women aged 20–74: United States, selected years 1960–1962



ing Medicare beneficiaries, ages 65+: US, 1992-2010 (Source: MCBS)

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
70.2	69.3	69.4	69.5	68.1	69.7	68.6	70.6	70.5	71.9	70.8	70.4	70.1	71.7	
43.5	41.3	41.3	39.9	38.4	38.4	38.1	38.3	39.1	40.0	37.8	37.6	37.2	38.8	
32.1	31.0	30.7	29.6	27.8	28.6	27.2	28.3	29.6	28.0	27.5	27.0	27.5	28.1	
30.1	29.5	29.5	29.2	27.7	28.1	27.0	26.9	28.0	27.7	27.2	26.1	26.4	26.9	
46.8	46.3	46.1	45.7	44.2	45.2	45.3	46.3	46.8	47.4	46.2	46.0	46.2	48.1	
69.2	68.2	68.3	68.5	67.0	68.8	67.9	69.8	69.8	71.1	70.0	69.8	69.4	70.9	
41.4	39.0	39.0	37.6	36.0	36.2	36.2	36.3	37.1	38.0	35.7	35.8	35.3	36.9	
30.2	29.2	28.7	27.9	26.0	26.8	25.3	26.7	27.8	26.3	25.7	25.5	26.1	26.6	
28.6	27.8	28.0	27.8	26.2	26.6	25.6	26.5	26.1	25.8	24.9	25.3	25.7		
45.1	44.4	44.3	43.9	42.3	43.5	43.8	44.8	45.3	45.8	44.6	44.7	44.8	46.8	
88.0	90.3	90.7	91.5	92.2	93.3	87.8	90.4	88.9	92.4	90.1	91.2	94.2	91.2	
84.6	86.9	88.2	88.4	89.4	92.3	86.9	87.5	90.5	91.5	91.1	88.3	92.6	90.2	
62.9	65.2	64.9	62.9	64.4	68.3	68.1	62.6	69.0	68.5	66.2	66.8	72.1	67.3	
62.0	64.3	60.1	59.2	63.5	60.5	61.7	58.0	64.1	65.9	59.5	59.1	63.5	59.5	
81.9	85.0	85.8	86.8	90.2	88.7	87.5	85.9	84.8	89.0	85.9	90.0	91.2	88.3	

Is there a better way to  
 Enable discovery  
 Gain insight  
 Tell compelling stories  
 ???

# SEARCH FOR THE TOOL

[HTTP://SELECTION.DATAVISUALIZATION.CH/](http://selection.datavisualization.ch/)

+ DATAVISUALIZATION.CH SELECTED TOOLS

Search...

All

Maps

Charts

Data

Color

Code? ✓



## Arbor.js

A library of force-directed layout algorithms plus abstractions for graph organization and refresh handling.



## CartoDB

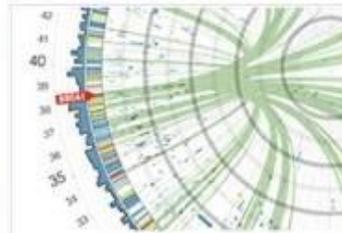
A web service for mapping, analyzing and building applications with data.

<http://cartodb.com>



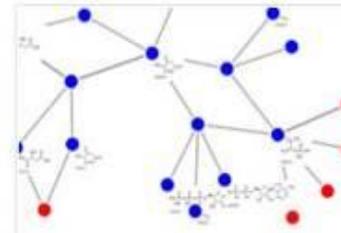
## Chroma.js

Interactive color space explorer that allows to preview a set of linear interpolated equidistant colors.



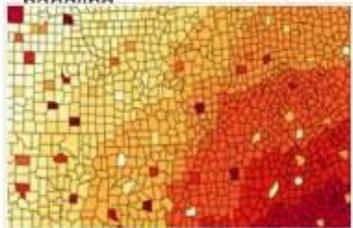
## Circos

A software package for visualizing data in a circular layout.



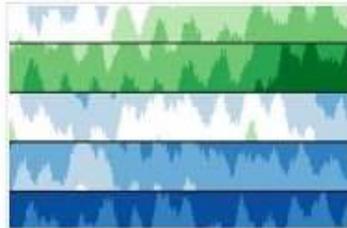
## Cola.js

A library for arranging networks using constraint-based optimization techniques.



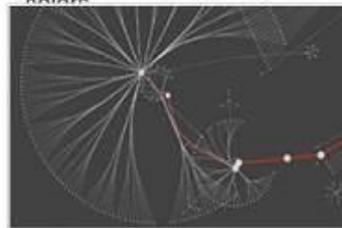
## ColorBrewer

A web tool for selecting colors for maps.



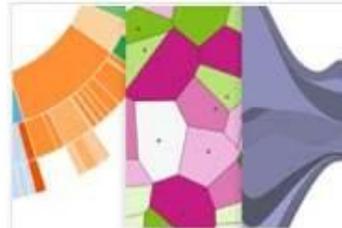
## Cubism.js

A library for creating interactive time series and horizon graphs based on D3.js



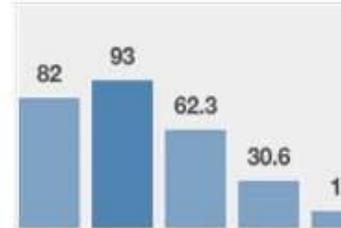
## Cytoscape

An application for visualizing complex networks and integrating these with any type of attribute data.



## D3.js

An small, flexible and efficient library to create and manipulate interactive documents based on data.

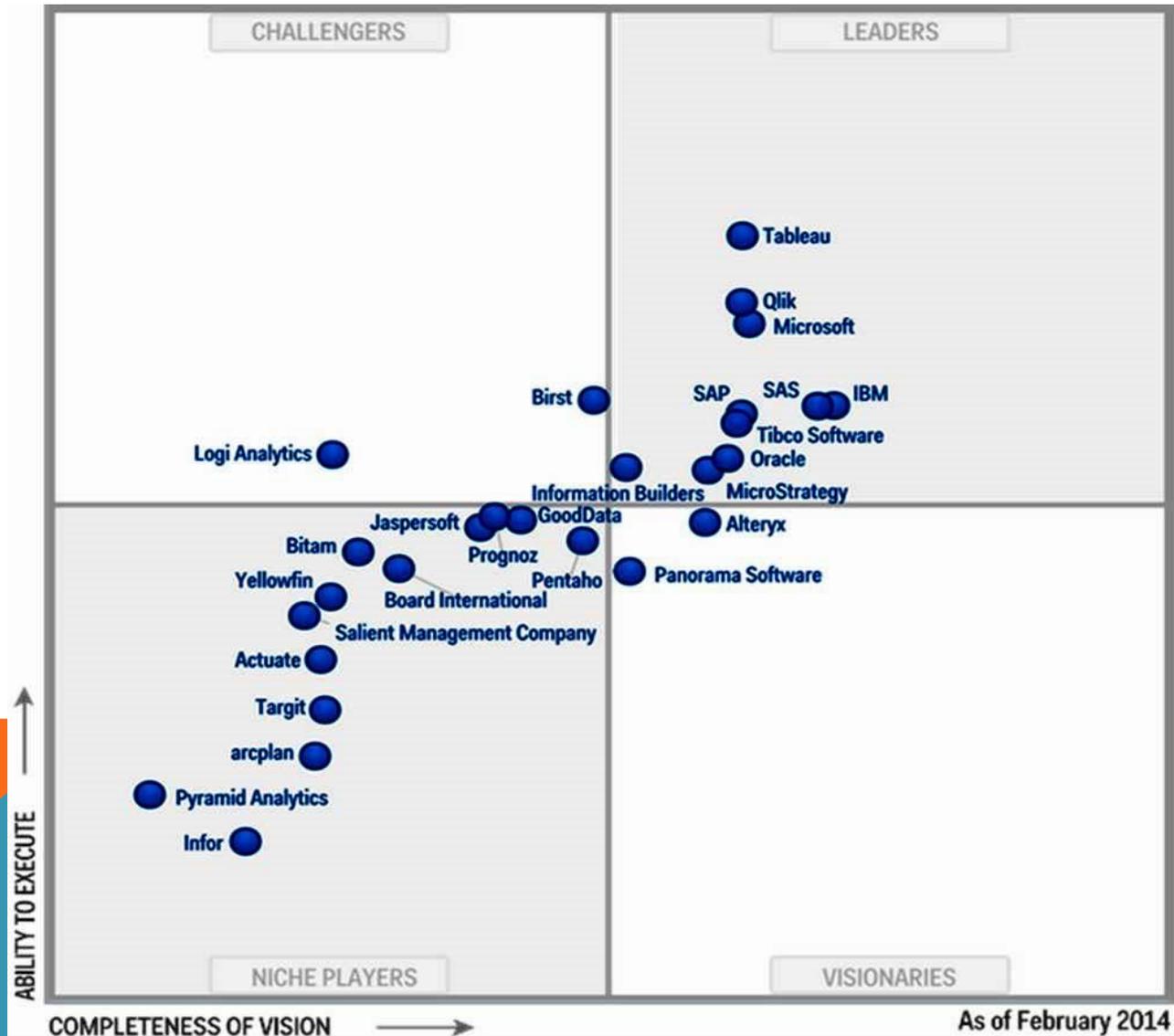


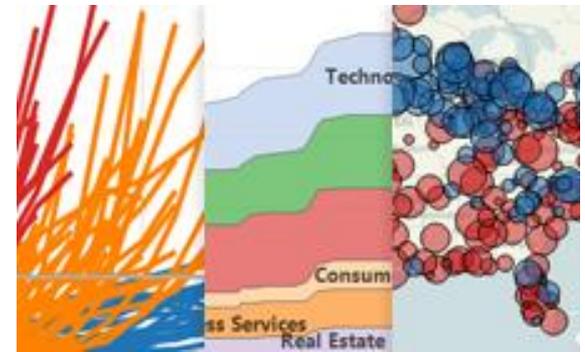
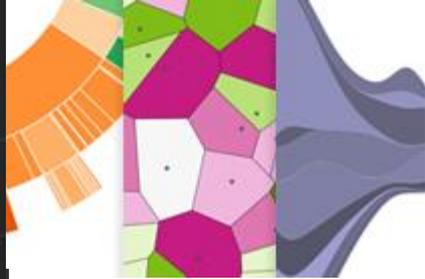
## Dance.js

A simple data-driven visualization framework based on Data.js and Underscore.js

# SEARCH FOR THE TOOL

## GARTNER'S MAGIC QUADRANT FOR BI PLATFORMS





# SEARCH FOR TOOLS

*A lot to choose from – which one fits NCHS the best?*

## Tableau

### Considerations and constraints:

- Programming vs drag and drop
- Data integration vs proprietary format
- Learning curve and user friendliness
- IT platform requirements
- Costs (end of fiscal year – no spending)
- CDC approved software

# TABLEAU ARCHITECTURE

Tableau Desktop  
Professional: design

## Tableau Native Query Technology

- Interactive queries to the database
- Leverages IT policies for security
- Avoids data silos
- Ensures fresh data results
- Values the EDW strategy

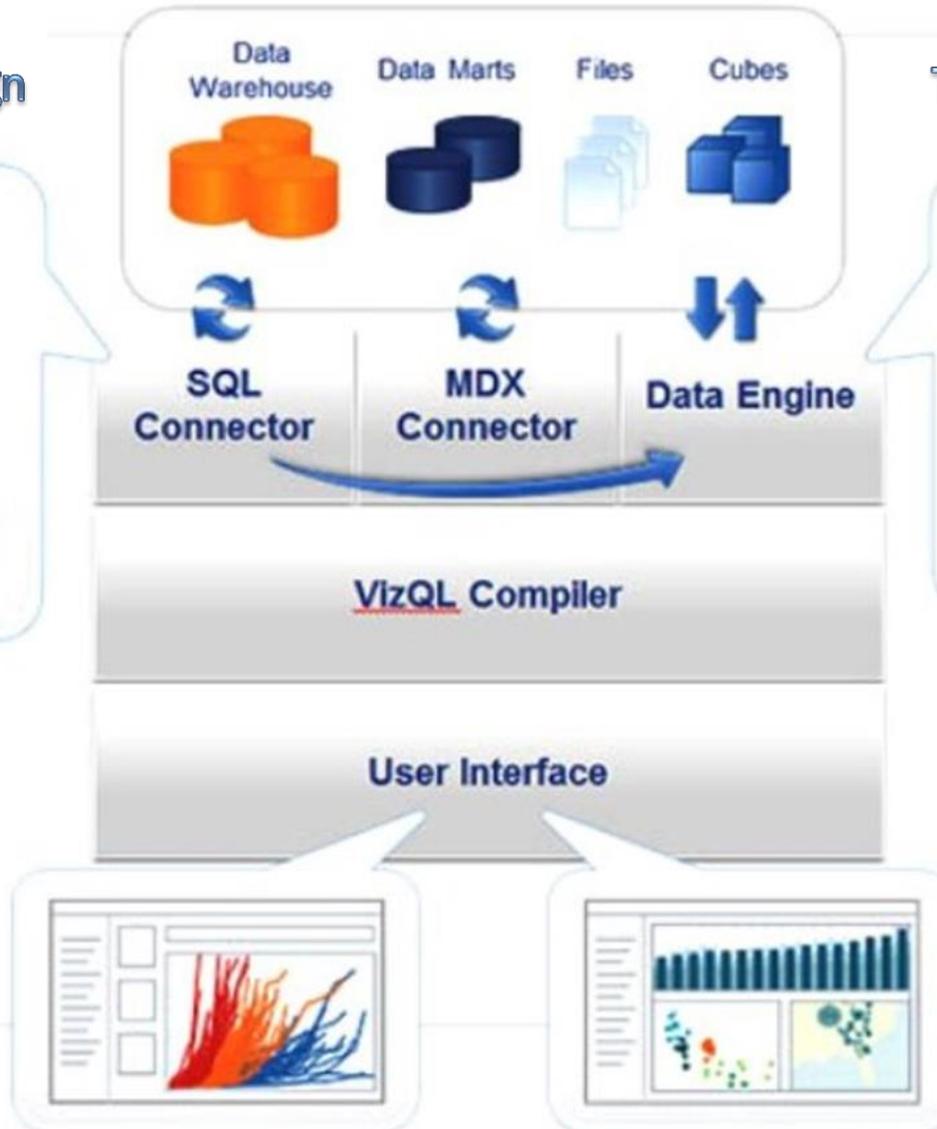
Tableau Server  
Trial: publish on web

## Tableau Data Engine Technology

- In-memory
- Column store
- Highly compressed
- Optimized API specific for Tableau
- 64-bit (32-bit version as well)

### Use this option:

- When the source DB query performance is slow
- To offload iterative query workload from the source DB
- To work offline from the network
- To keep an archive of the data



# VM

Virtual Server  
Informatics  
Innovation Lab

# PREPARE THE DATABASE

## In a file:

- Tableau data extract, workbook
- Microsoft access, excel, text

## On a server:

- Tableau server
- MS SQL, MySQL, SAP, Oracle
- Hadoop, Redshift, Google BigQuery, ODBC
- 30+ different data formats

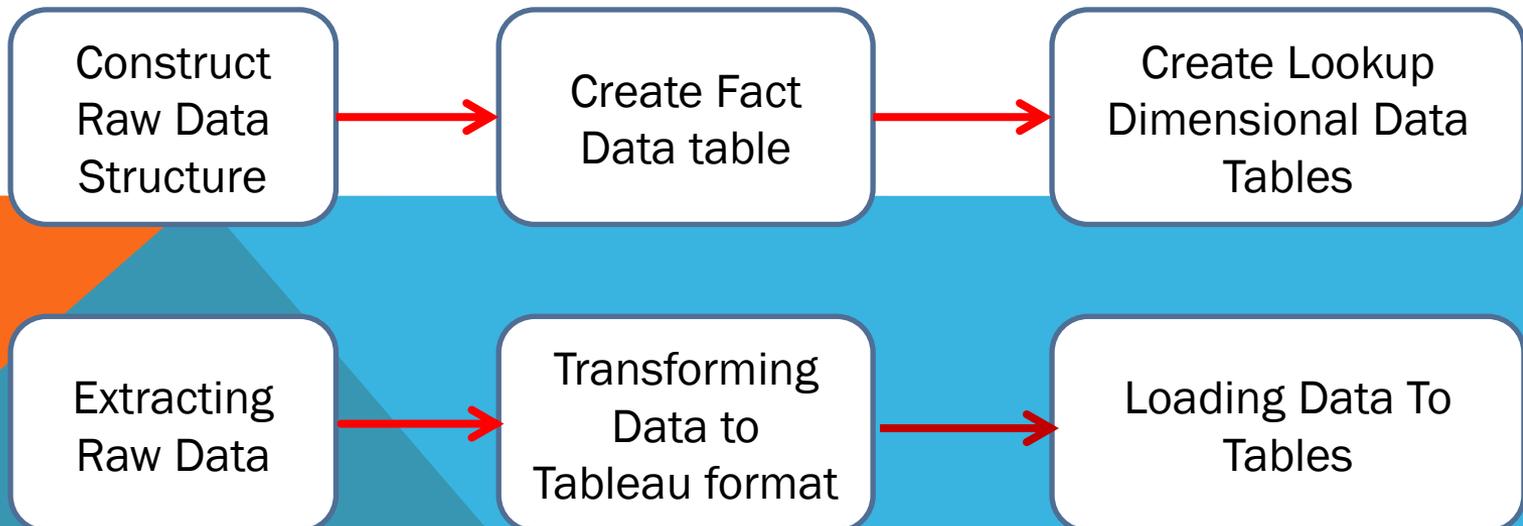
## Flexibility to Choose How You Work With Data



Direct connect to fast databases



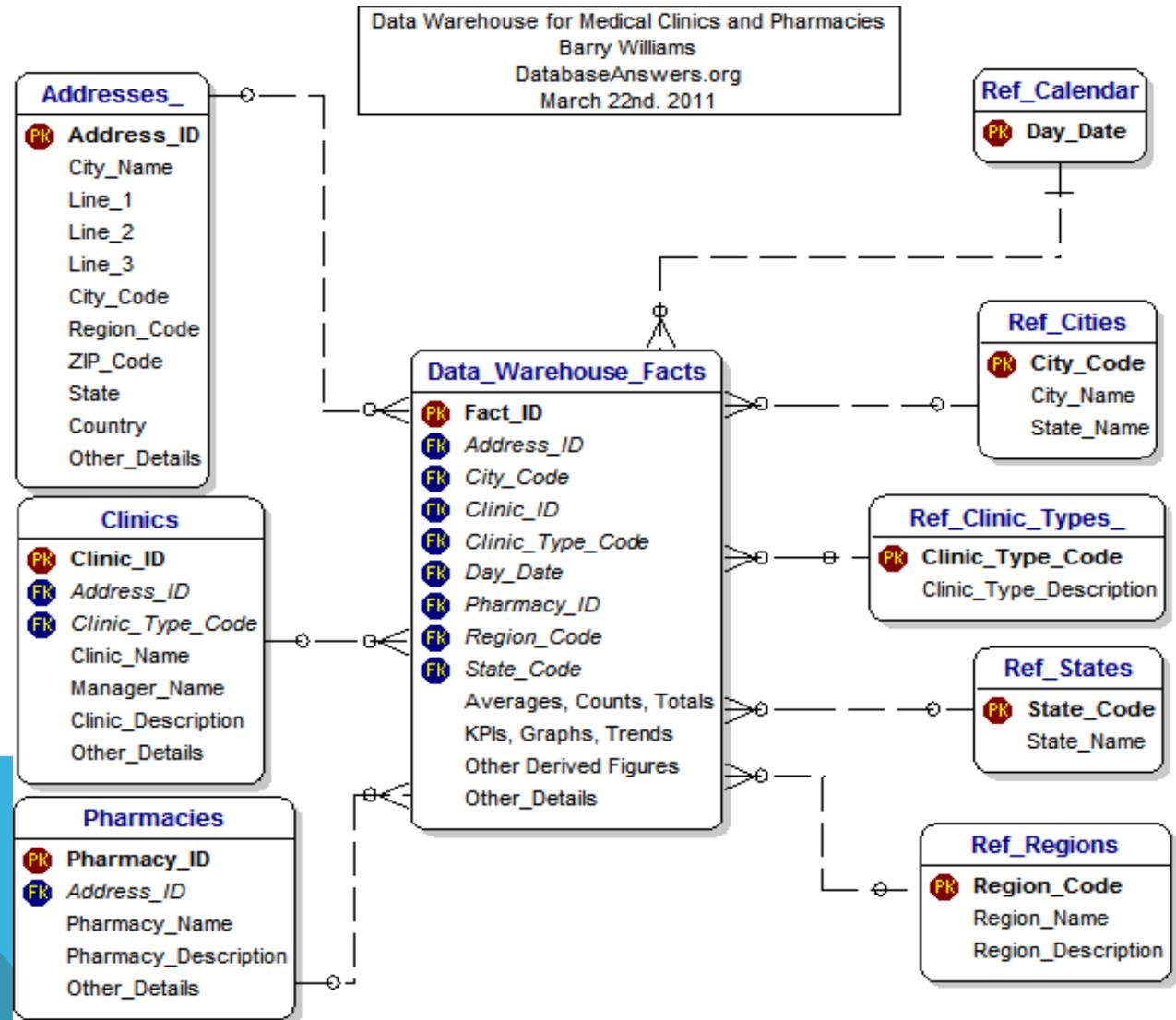
In-memory accelerates slow databases



# EXTRACTING, TRANSFORMING AND LOADING DATA

ETL process:

- Extract fields and load data files into raw data table
- Run SQL script to join raw data with fact table and dimension table
- Add columns and update if necessary
- Perform Tableau data extraction



# CONNECTING DATA TO TABLEAU

The screenshot shows the Tableau Desktop interface. At the top, the window title is "Tableau - Book12" and the menu bar includes "File", "Data", "Server", and "Help". The main view is titled "mortality9911r4+ (vitalstatistics)" and shows a connection to a MySQL server at "localhost" with the "vitalstatistics" database selected. A list of tables is visible on the left, including "combined\_popul...tality\_0811web", "combined\_popul...mortality\_9911", "combined\_popul...mortality\_total", "combined\_popul...race\_recod4", "combined\_popul...talracerecode4", "dim\_state\_coun...\_residence\_raw", "dim\_state\_county\_residence", "fact\_standardpop2", "hispanic\_origin", "hispanic\_recode9", and "icd10".

In the center, a diagram shows two data sources, "mortality9911r4" and "icd10\_recode113", connected by a Venn diagram icon. A "Join" dialog box is open, showing the "Inner" join type selected. The dialog also displays the join clause: "icd10 Recode113 id = icd10 Recode113 id (icd...".

Below the diagram, a "Copy" button is visible. At the bottom, a data table is displayed with the following columns: "icd10 Recode113 id (icd10 recode113)", "Icd10 Recode113 Desc", "Mort Id", "Race Ethnicity Recode4 Id", "State Alpha Code", "Death Month", "Death Year", and "Sex". The table contains four rows of data:

icd10 Recode113 id (icd10 recode113)	Icd10 Recode113 Desc	Mort Id	Race Ethnicity Recode4 Id	State Alpha Code	Death Month	Death Year	Sex
1	Salmonella infections (A01-A02)	39394325	3	AL	07	2000	M
1	Salmonella infections (A01-A02)	39420977	2	AZ	03	2000	F
1	Salmonella infections (A01-A02)	39487874	4	CA	01	2000	F
1	Salmonella infections (A01-A02)	39633107	2	CA	08	2000	F

Can join, blend multiple  
data sources

## Tableau Data Blending



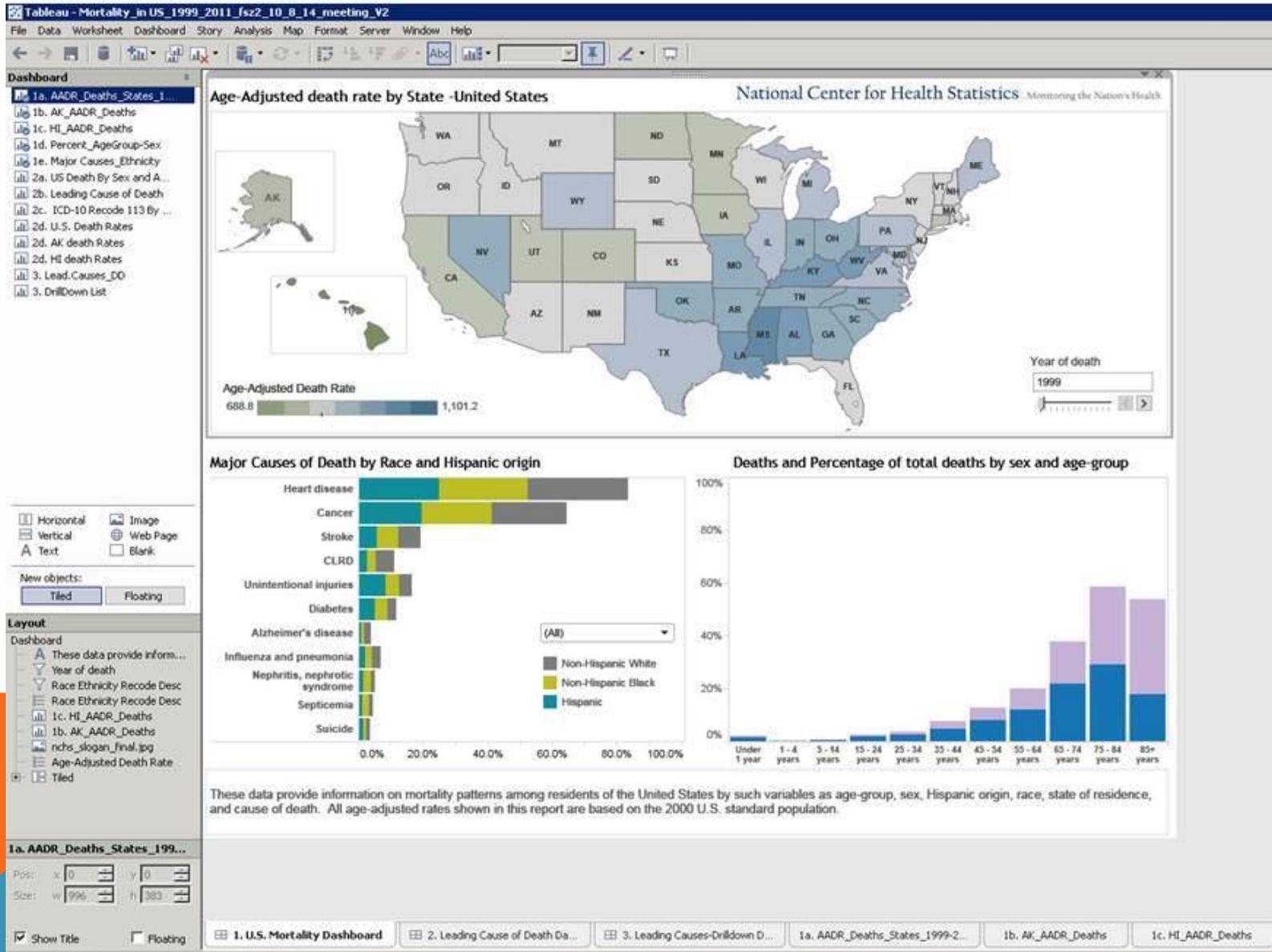
# DATA PREPARATION

The devil is in the details: behind “drag & drop”, thoughtful preparation has to happen behind the scene!

## Speak Tableau Language!

- Select and package the data for visual analytics and end users
- Interactivity: filters, parameters, sets, actions do NOT just happen
- Budget 1/3 of time for database building
- Team up database experts with analysts

# MORTALITY DASHBOARD: GETTING STARTED

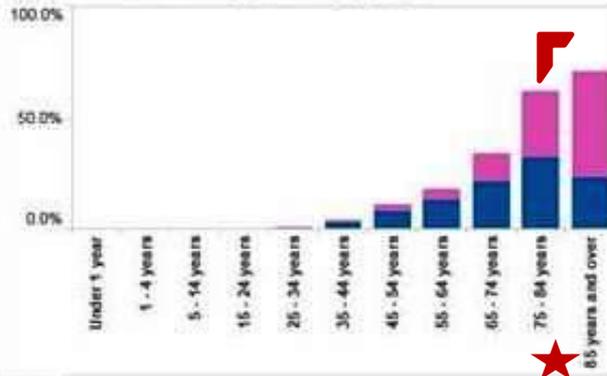


# TELLING STORY OF HEART DISEASES

Death Rates per 100,000 population



Percent of deaths by Sex and Age groups



2001 vs 2011:

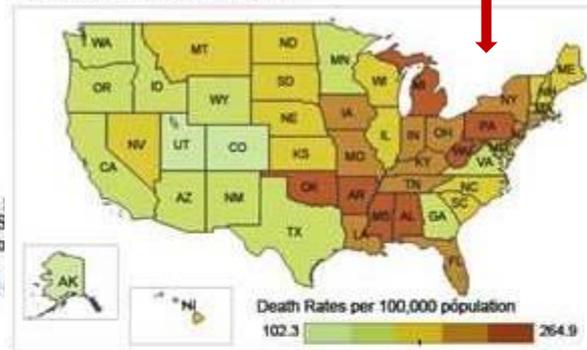
- ↓ See changes in death rates in states
- ★ See total number of deaths dropped
- ↙ See changes in age/sex groups

Leading Causes of Death



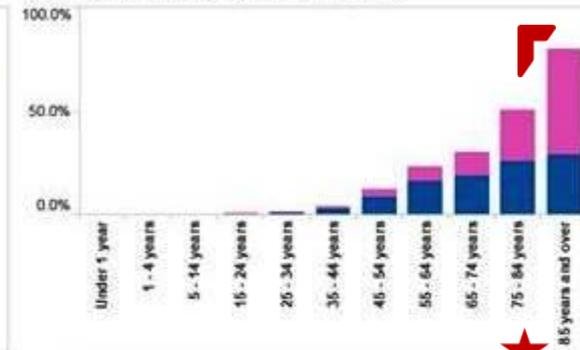
2001 Total number of deaths in 2001 for Diseases of heart, 700,142

Death Rates per 100,000 population

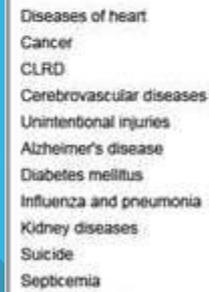


These data provide information on mortality patterns among residents of the United States and cause of National

Percent of deaths by Sex and Age groups



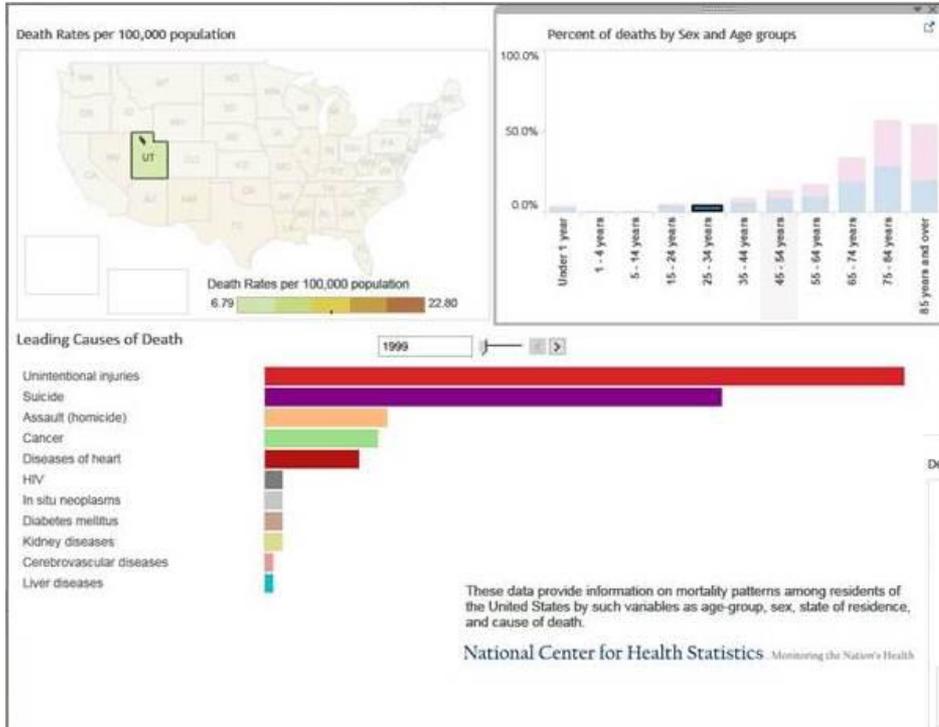
Leading Causes of Death



2011 Total number of deaths in 2011 for Diseases of heart, 596,574

These data provide information on mortality patterns among residents of

# COMPARE LEADING CAUSES FOR SELECTED STATES AND AGE GROUPS

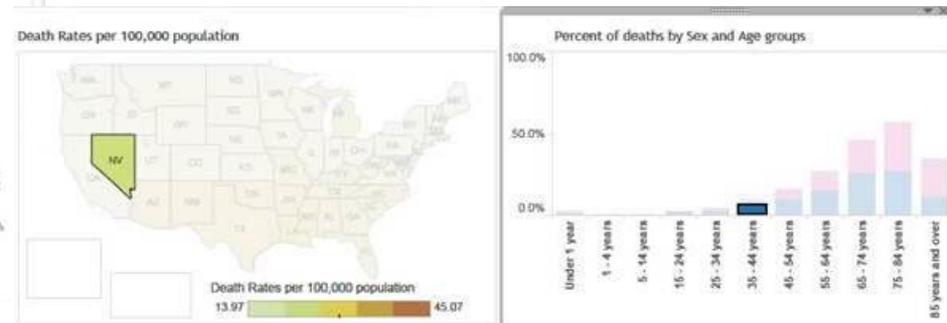


## Utah

- Unintentional injury
- Suicide
- Assault
- Cancer

## Nevada

- Unintentional injury
- Heart Disease
- Suicide
- Cancer



## Leading Causes of Death

- Unintentional injuries
- Diseases of heart
- Suicide
- Cancer
- HIV
- Liver diseases
- Assault (homicide)
- Cerebrovascular diseases
- Diabetes mellitus
- Septicemia
- In situ neoplasms
- Aortic aneurysm and dissection
- Influenza and pneumonia
- CLRD
- Kidney diseases
- Essential hypertension
- Pneumonitis due to solids and liquids

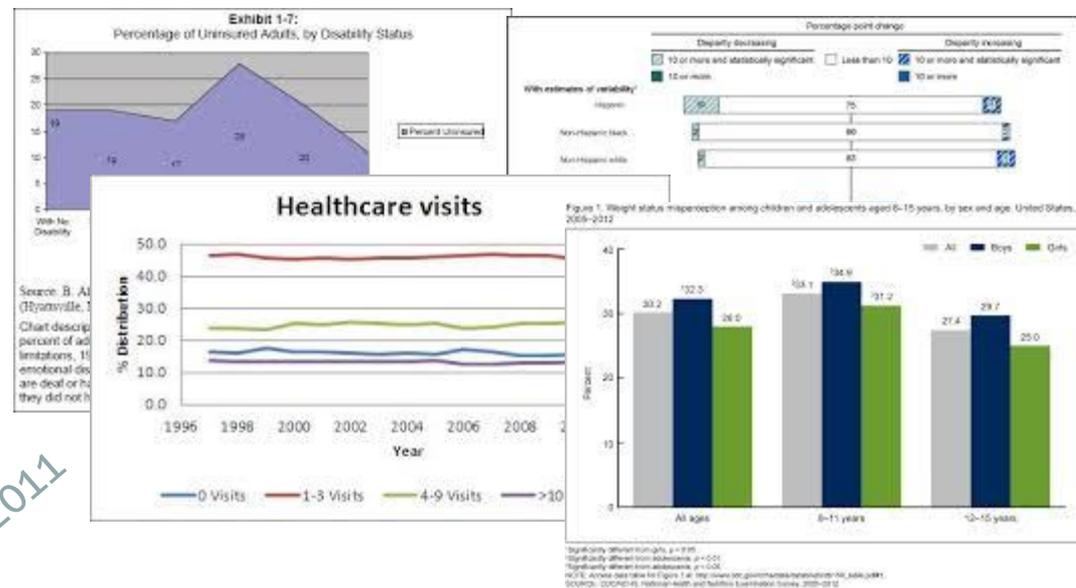
These data provide information on mortality patterns among residents of the United States by such variables as age-group, sex, state of residence, and cause of death.

National Center for Health Statistics - Monitoring the Nation's Health



# DASHBOARD

Mortality in United States, 1999-2011



## Flexible, expressive, Interactive:

- Combine multiple data sources in one dashboard – easy to find info
- Dynamic: allow filtering, highlighting, and drilling down
- Let end users in driver's seat – fast creation of charts of interest
- Automated updates with new data
- Easy sharing with support for existing security

# CHALLENGES IN USING TABLEAU

- **Remote connection: network connectivity and latency**
  - **Original configuration: 2 cores, 8 GB Mem, 16 GB Disk**
  - **12+ hours data extraction: 13-year data, 5 fields, 35 million records**
  - **Later configuration: 6 cores, 16 GB Mem, and 110 GB Disk.**
  - **Functional: slow when extracting data (overnight job)**
  - **Tableau cannot handle complex survey – need R integration**
  - **Tableau cannot read SAS data**
- 



# MOVING FORWARD

R Shiny and SAS Visual Analytic

## Superhighway for Innovation

**We learned to speak Tableau in 2 months:**

- **Tableau Story board for Health US**
- **Tableau integration with R**
- **Build templates in SAS VA for NHIS and NHANES (complex survey data)**
- **Test R/Shiny on survey data**