Let’s Talk Analytics

Role of Data Analysts in NSHA

Steven Carrigan, Manager, Performance and Analytics

May 30, 2019

Please be advised that we are currently in a controlled vendor environment for the

One Person One Record project.

Please refrain from questions or discussion related to the

One Person One Record project.

Informatics…

“Utilizes health information and health care technology to enable patients to receive best treatment and best outcome possible.”

Clinical Informatics…

"is the application of informatics and information technology to deliver health care."

AMIA. (2017, January 13). Retrieved from https://www.amia.org/applications-infomatics/clinical-informatics

Analytics…

“is the discovery, interpretation, and communication of meaningful patterns in data.”

“relies on the simultaneous application of analysis, statistics, computer programming and operations research to quantify performance.”

Objectives

At the conclusion of this activity, participants will be able to…

- Identify what knowledge and skills health care providers will need to use information now and in the future.
- Prepare health care providers by introducing them to concepts and local experiences in Informatics.
- Acquire knowledge to remain current with new trends, terminology , studies, data and breaking news.
- Cooperate with a network of colleagues establishing connections and leaders that will provide assistance and advice for business issues, as well as for best-practice and knowledge sharing.

Conflict of Interest Declaration

- I do not have an affiliation (financial or otherwise) with a pharmaceutical, medical device, health care informatics organization, or other for-profit funder of this program.

NSHA Role of Analytics

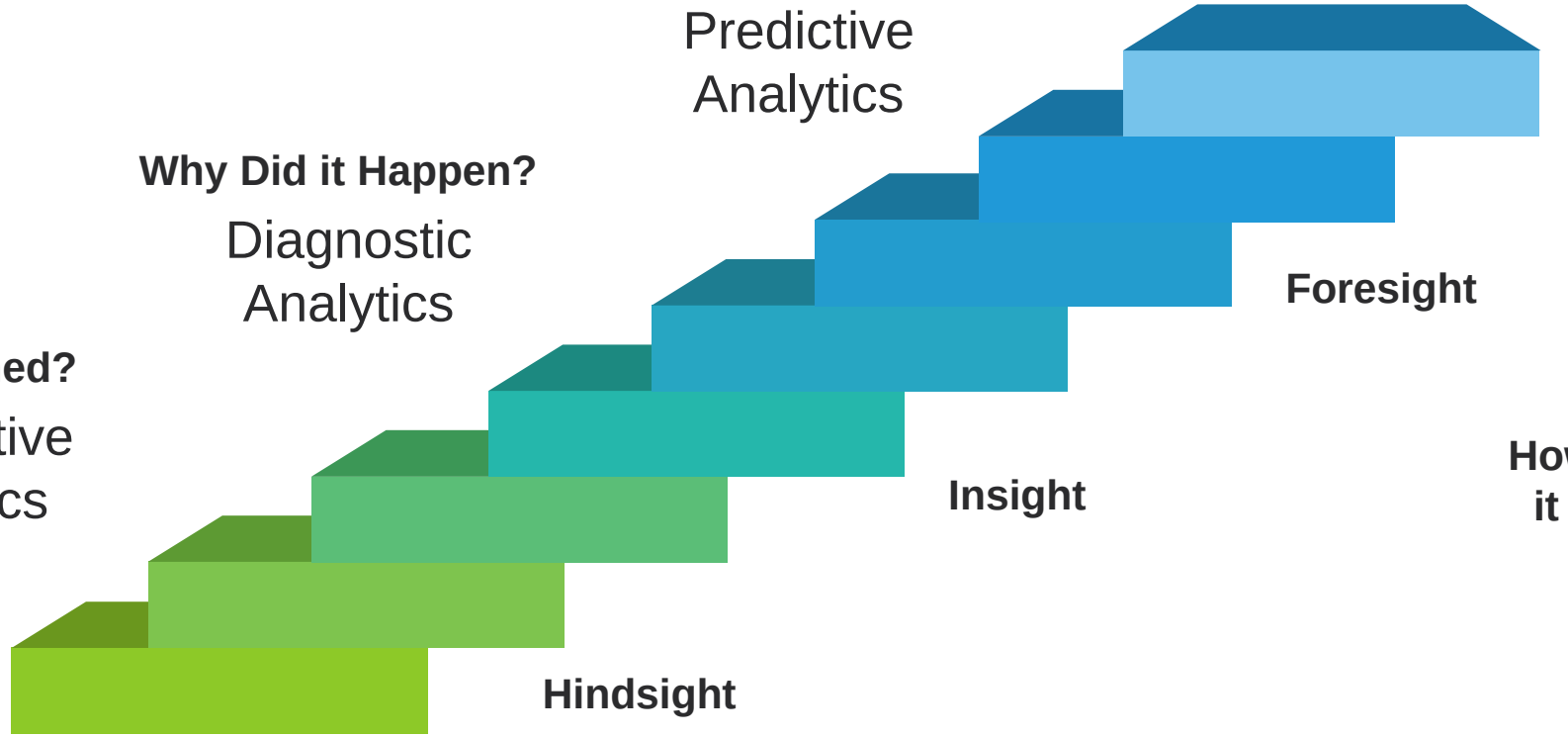
The specific objectives of this session are to help you understand the role of Analysts and Analytics in NSHA

01 Performance and Analytics

02 Analytics/Data Science

03 Applied Analytics - Case Studies

Steps of the Analytics Continuum



Performance & Analytics Team

A team of 25 analysts spread across the province, endeavoring to provide data to drive decision making throughout NSHA

"Supporting evidence and data based decision making"

The team is made up of a mix of roles with team members having varied backgrounds and skill sets to support a wide range of analytical services and projects.

- Senior Decision Support Analysts
- Data Analysts
- Decision Support Analysts
- Research and Statistics Officers
- Health Records Analysts
- MIS Statistical Coordinators

What Do We Do

Performance & Analytics Team

Data Extraction and Staging

Data extraction, aggregation and data mining methods organize the data and make it possible to identify patterns and relationships in it that would not otherwise be visible.

Inferential Analytics

Examines data and relationships to answer the question “Why did it happen?”, or “What might happen?”.

Data Linkage

Allows related information from one data source to be linked to information from another data source. Using the linked data makes it possible to gain a more comprehensive understanding.

Performance Measurement

Regular measurement of outcomes and results, which generates reliable data on access, efficiency and outcomes for NSHA programs

Descriptive Analytics

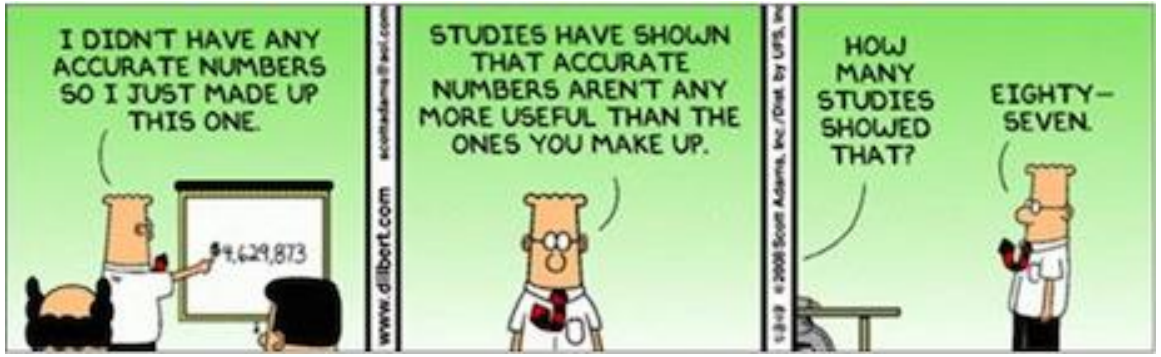
Summary of historical data to yield useful information and possibly prepare the data for further analysis. Reporting and data visualization may be applied to yield more insight.

Applied Analytics

Utilizing advanced statistical methods and innovative tools to create iterative, scalable and operational solutions, applying theory to the practice of health system planning and management.

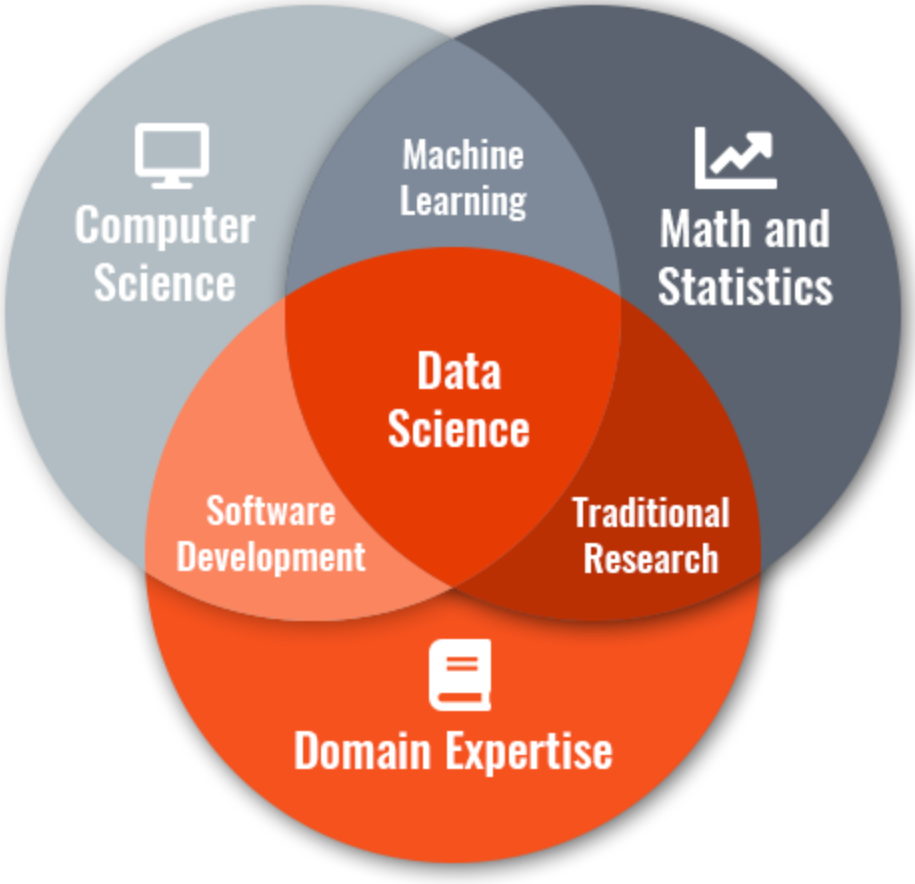
What We Don’t Do

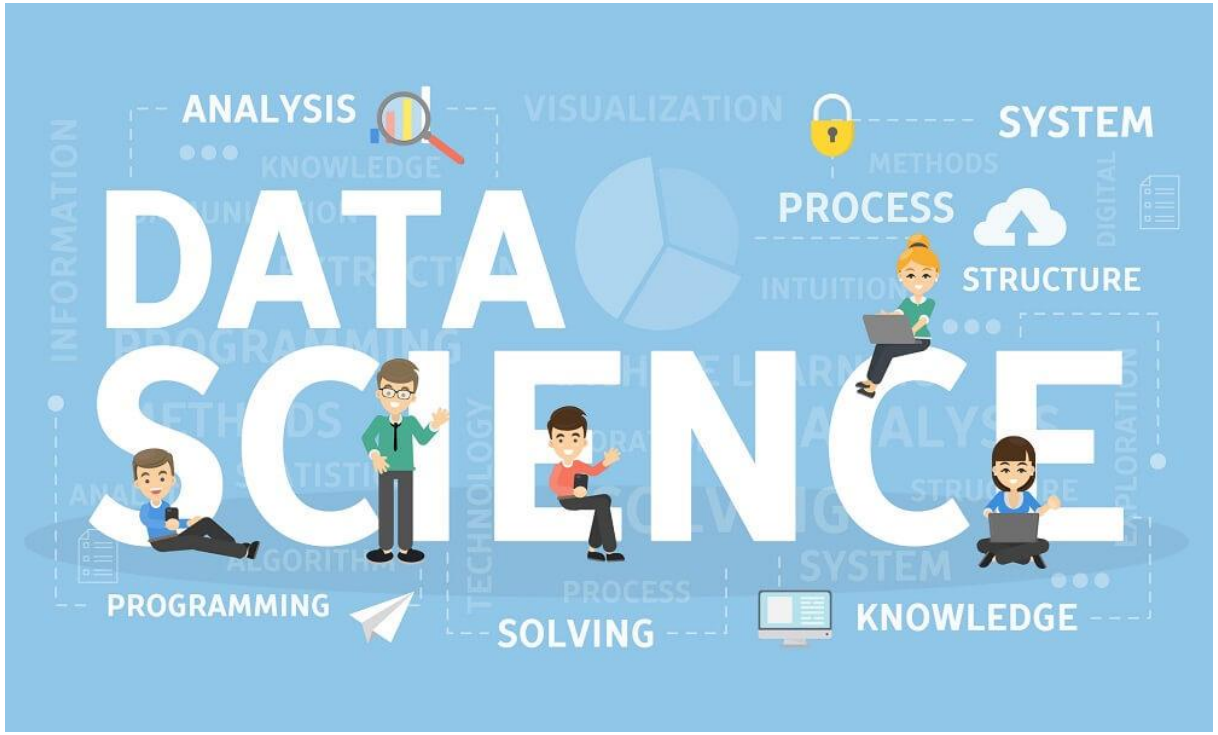
“After analyzing all your data, I think we can safely say that none of it is useful.”



What is Data Science

and, what it isn't





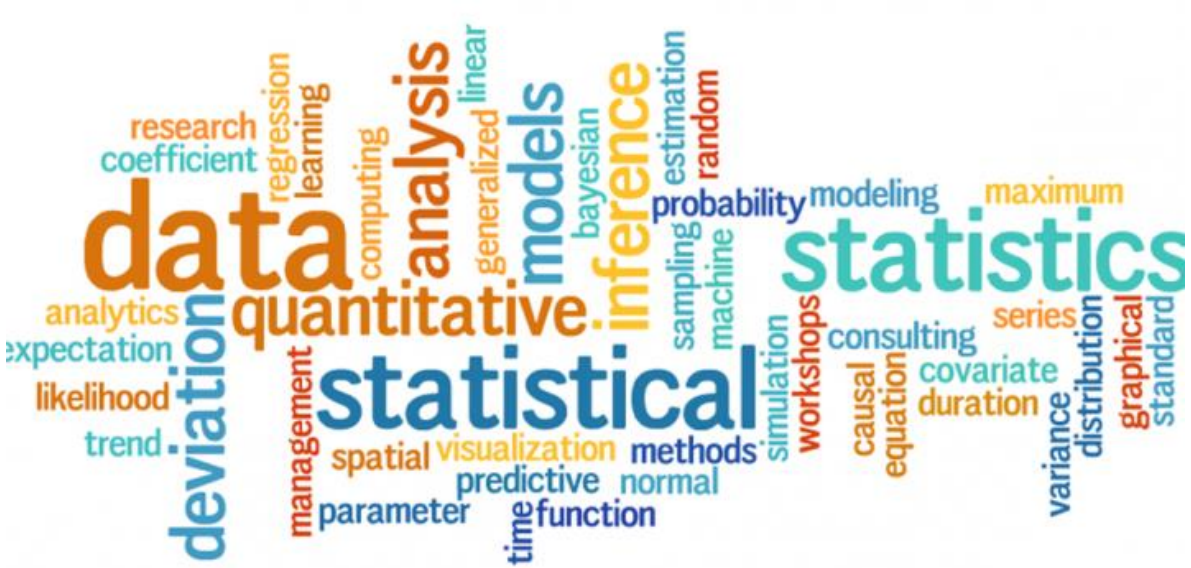
Analytics

Telling the Story in the Data

- Transforming data into insight, rapid turnaround, often reactive.
- Analysts are data storytellers.
- Summarizing interesting facts and using data for inspiration.
- Domain knowledge and understanding/translation of analytical/statistical concepts to support decision-makers.

Statistics

Determining which data and conclusions are trustworthy

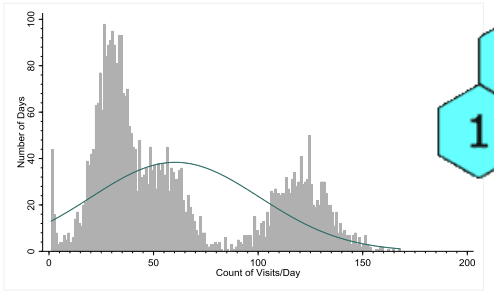


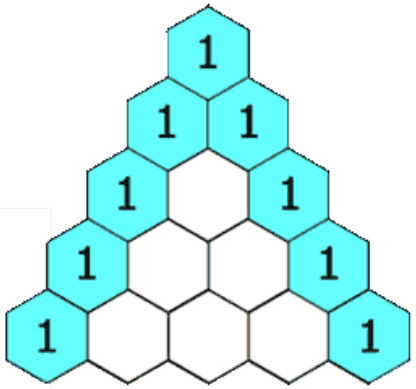
- Statistics allows us to come to conclusions beyond our data (inferences).
- Ensuring the methods applied are right for the problem and inferences are valid from the information at hand.
- People add the new information to their own pre-conceptions of the data (bias).

Statistics

Identifying and Understanding Patterns

- Understanding the patterns (distributions) and relationships in our data.
- Using these patterns to guide decisions and understand the past and predict the future
- Sorting out the signal from the noise (meaningful patterns)



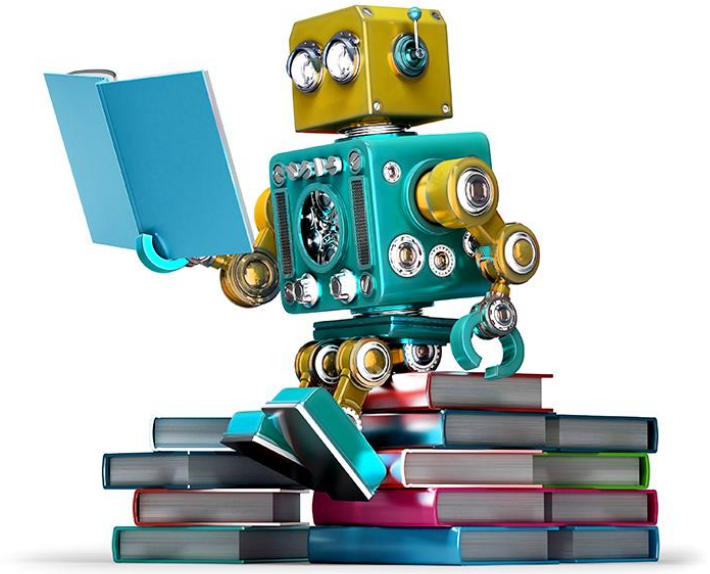


Machine Learning

Automation and Performance

- Machine learning is a method of analytics that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

- Performance means reliable, scalable, and easy-to-maintain models that perform well in practice.



Wide versus Deep

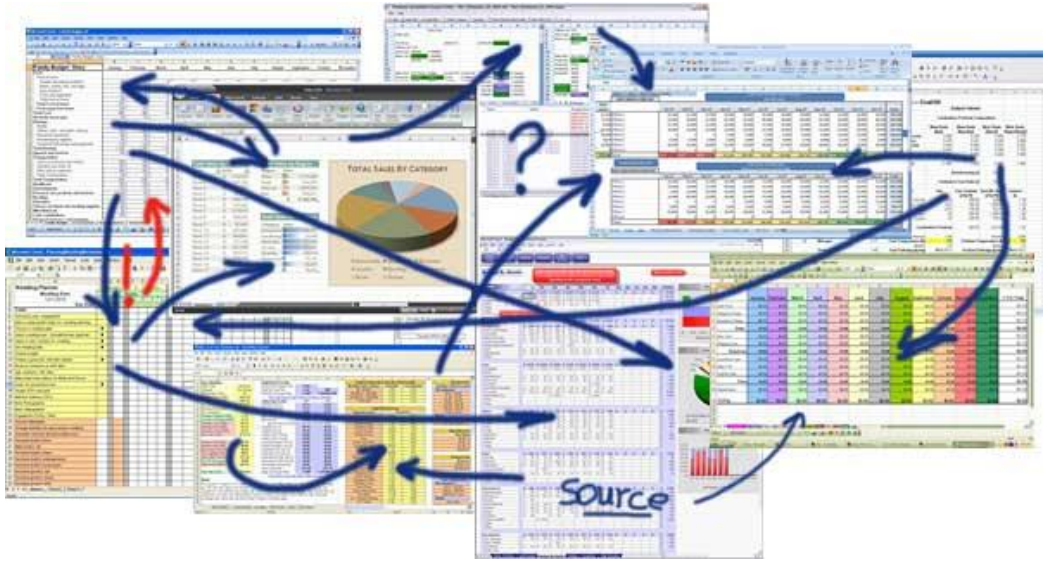
Avoiding the “rabbit hole”

- The previous two domains of analytics both provide high-effort solutions to specific problems. If the problems they tackle aren’t worth solving, you end up wasting time and resources.
- Statistics and machine learning are narrow-and-deep pursuits — the shape of a rabbit hole, incidentally — it’s important to tackle problems that deserve the effort.

Applying Data and Theory to Problems

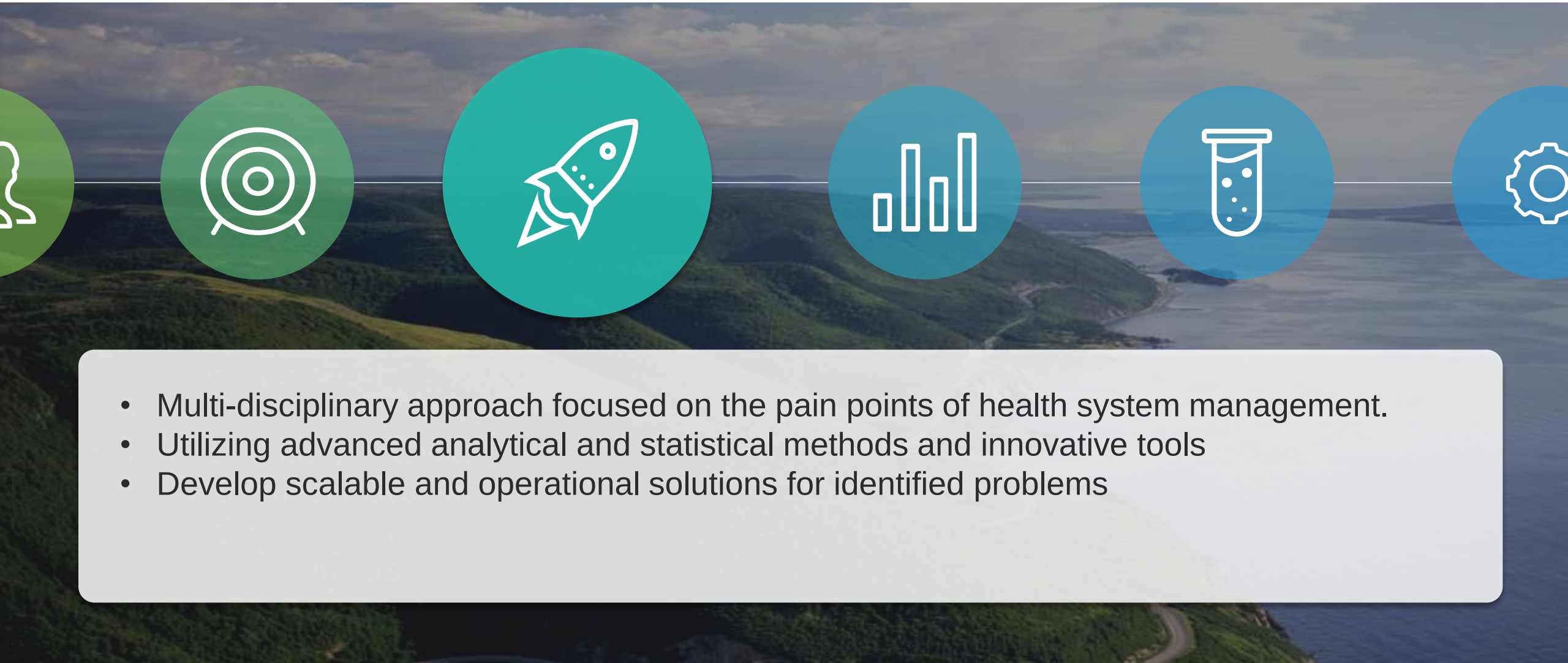
Transform Patient Data into Intelligent Action

- As an organization we're constantly producing data; numbers, reports, trend lines, graphs, spreadsheets and more.
- But, do we know if all this data leads to better decisions?



Applied Analytics

Applying theory to the practice of health system management



- Multi-disciplinary approach focused on the pain points of health system management.
- Utilizing advanced analytical and statistical methods and innovative tools
- Develop scalable and operational solutions for identified problems

Applied Analytics Methodology | Transform Patient Data into Intelligent Action

Problem → **Solution** → **Value**

Data

- Data Sources
- Apps and Web
- Sensors and Devices

Applied Analytics

Data Analysis
- Information Exploration
- Data Analysis

Modelling and Validation
- Exploratory Models
- Time-Series Analysis
- Machine Learning

Action

- People
- Reporting
- Automated Systems

Data Requirement

e.g. emergency department visits, surgical cases, outpatient visits, midnight stays history

Modeling

Use multiple models to produce forecasts/projections

Improved Operations

Potential Short-Term Forecasting	Potential Long Term Forecasting
Utlization	Demographics
Weekly cycles	Industry trends
Daily actuals	Policy trends

nova scotia health authority

Analytical Case Studies

The specific objectives of this session are to help you understand the role of Performance and Analytics in NSHA and specific analyst roles

01 Utilization Forecasts

02 Utilization - Structural Equation Modelling

03 Population Based Projections

04 Predicting Patient Flow

Applied Analytics | Utilization Volume Modelling and Forecasting

Business Problem
Lack of forecasts of health system utilization to support planning

Solution
Patient volume forecasting based on time based model

Business Value
Enhanced forecasts leading to operational improvements and improved patient experiences

Data

Data Sources

Inpatient and ED administrative data for key utilization metrics

Applied Analytics

Data Analysis

Information Exploration

Data Analysis

Modelling and Validation

Polynomial Regression Model

Action

Dissemination to leaders

Integration into Regular Reporting

Testing of results

Data Requirement

Patient Volume History

e.g. emergency department visits, surgical cases, outpatient visits, etc.

Modeling

Examination of multiple models to produce forecasts

Time Series

Polynomial Regression

Linear Regression

Autoregressive–moving-average (ARMA) models

Improved Forecasting

Potential Short-Term Forecasting

Scheduling

Weekly cycles

Daily actuals

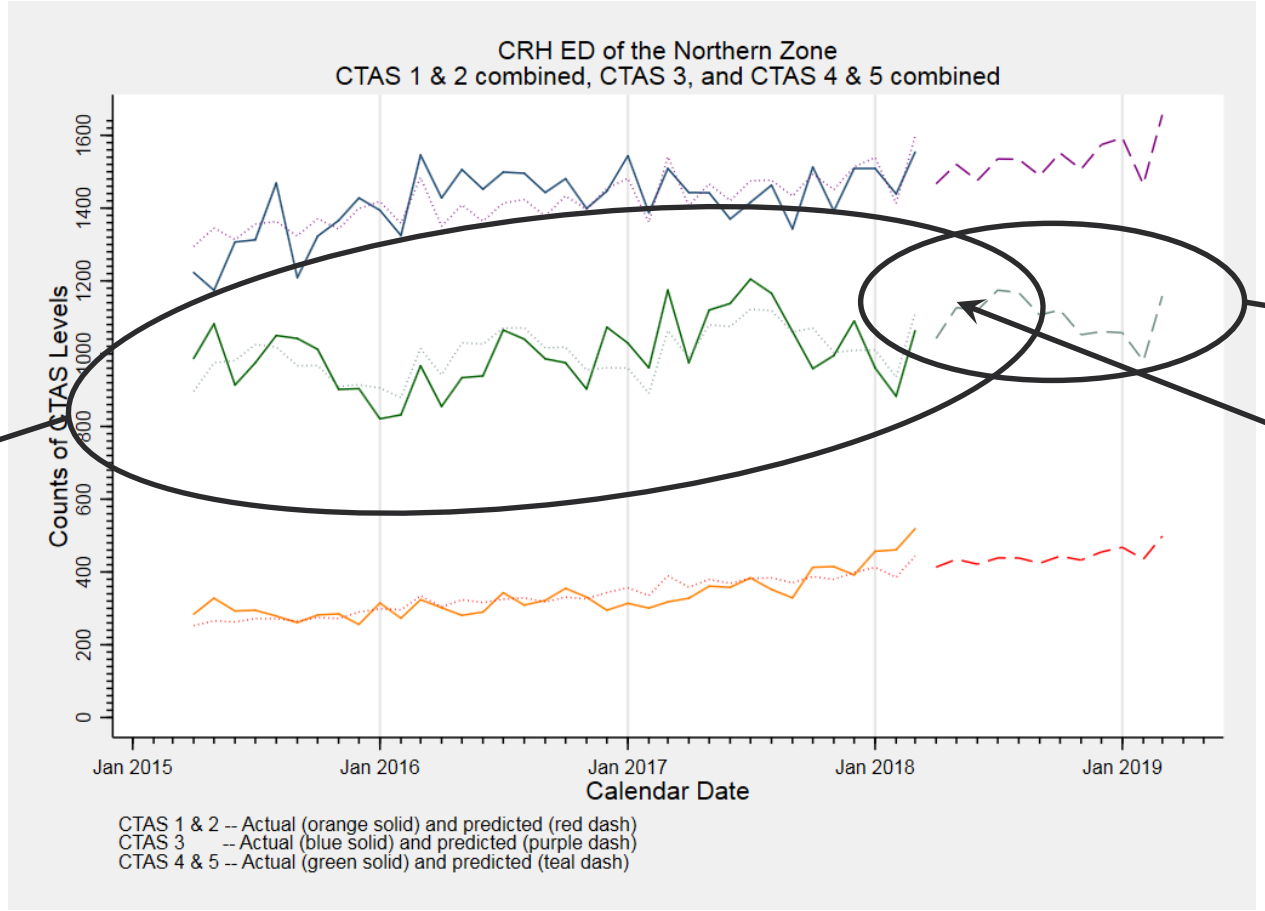
Potential Long Term Forecasting

Demographics

Utilization trends

Utilization Modelling and Forecasting

Understanding the Past to Predict the Future



Forecasting

testing

Modelling

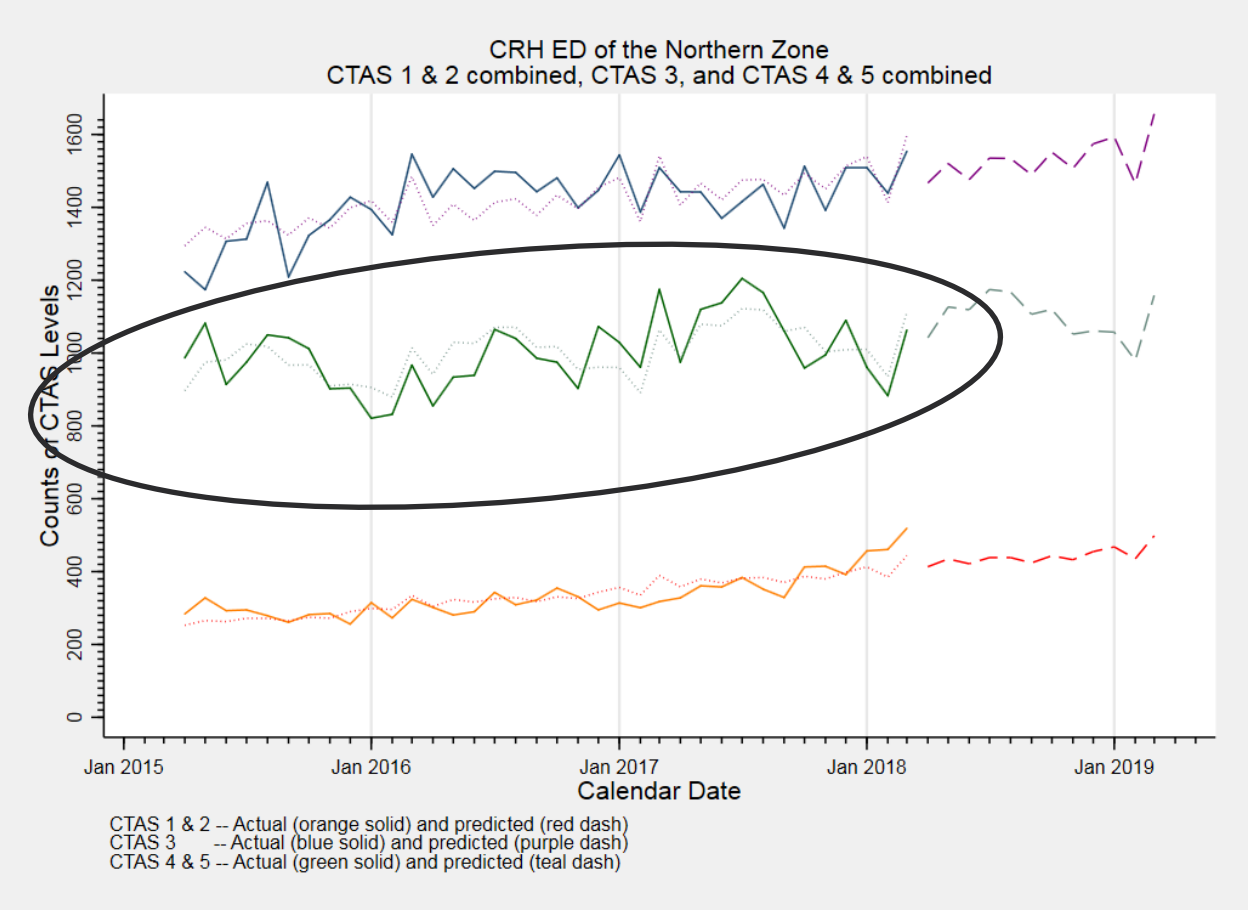
patterns
cause and effects
associations

- Forecasting is key, but the quality (utility) of the forecast depends on the quality of the model
- The quality of the model depends on the choice of variables, and the quality of measurement

Modelling - Understanding the Past

The quality of the model depends on the choice of variables, and the quality of measurement

- What is modeled?
  - Dependent Variable (visits, occupancy, admissions, etc.)
- How is it modeled?
  - Independent Variable or Predictors
  - prior pattern of the DV -- the future will look like the past
  - this is “time”
  - It works reasonably well



Modelling

The Process

Problem Statement

Variables

a, b, c, X, Y, Z

d, e, f, g, h, i
j, k, l, m, n
O, P, Q, R, S
T, U, V, W

η, θ, λ, ξ, π
α, β, γ, δ, ε, ζ,

Choose the Analysis

multiple regression
logistic regression
Poisson regression
ARIMA

Analyze

oh no ... too complicated
get rid of variables

Present

reduced to a 1 page
executive summary

Modelling

The Actual Analysis

The analysis is all done in Stata – it is quite straightforward currently

This little bit of code does all the work of modelling (and forecasting).

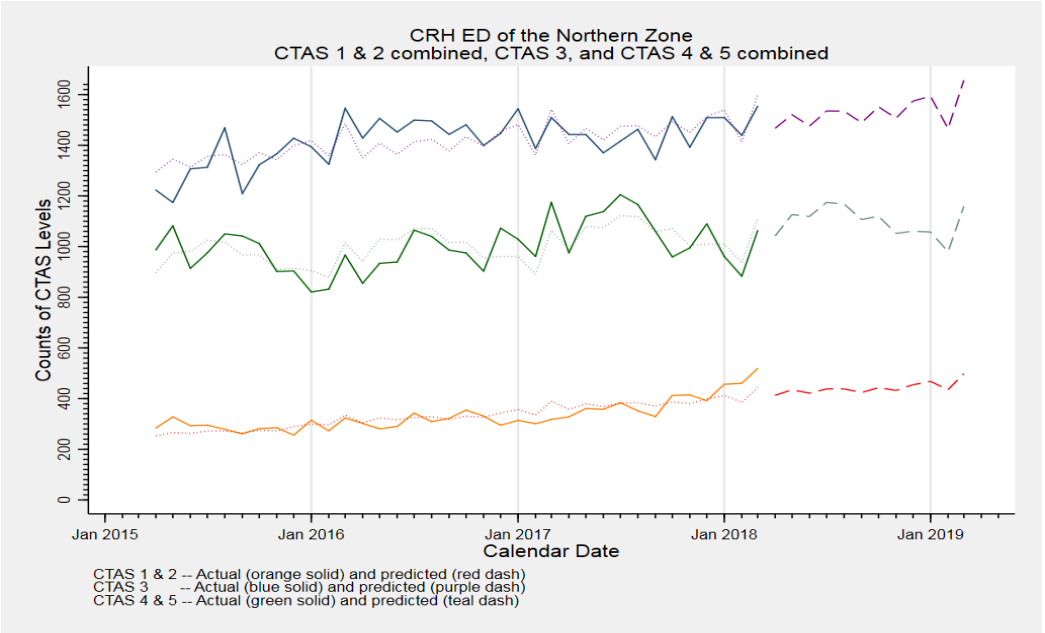
```
regress ED_Board                                                    ///
        DayNum_1 DayNum_2 DayNum_3 DayNum_4 DayNum_5                ///
        FM_1 FM_2 FM_3 FM_4                                         ///
        WkDay_1 WkDay_2 WkDay_3 WkDay_4 WkDay_5 WkDay_6             ///
        , cformat(%9.3f) pformat(%5.3f) sformat(%6.3f)
  local Fit  : display %6.3f e(r2)
  local FValue  : display %6.3f e(F)
  scalar PValue = 1- F(e(df_m),e(df_r),e(F))
  local pValue : display %6.3f PValue
  local Sample : display  e(N)
  display "Boarding: "
  display "R2 = `Fit' (F=`FValue', p<`pValue'), based on `Sample' points. "
  predict Expected
  predict StdErr, stdp
  predict ForErr, stdf
  gen Error  = temp - Expected
  gen LCI95 = Expected + invnormal(.025)*(ForErr )
  gen UCI95 = Expected - invnormal(.025)*(ForErr )
```

```
      Source |       SS           df       MS      Number of obs   =     1,249
-------------+----------------------------------   F(15, 1233)     =     65.54
       Model |  3214.09803        15  214.273202   Prob > F        =    0.0000
    Residual |  4031.01326     1,233  3.26927272   R-squared       =    0.4436
-------------+----------------------------------   Adj R-squared   =    0.4369
       Total |  7245.11129     1,248  5.80537764   Root MSE        =    1.8081

------------------------------------------------------------------------------
  ED_Boarded |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    DayNum_1 |       0.003      0.001     6.111   0.000         0.002       0.005
    DayNum_2 |       0.000      0.000     8.706   0.000         0.000       0.000
    DayNum_3 |       0.000      0.000     4.451   0.000         0.000       0.000
    DayNum_4 |      -0.000      0.000    -6.328   0.000        -0.000      -0.000
    DayNum_5 |      -0.000      0.000    -5.492   0.000        -0.000      -0.000
        FM_1 |       0.060      0.052     1.162   0.246        -0.041       0.162
        FM_2 |       0.003      0.016     0.161   0.872        -0.028       0.033
        FM_3 |      -0.002      0.003    -0.695   0.487        -0.008       0.004
        FM_4 |       0.000      0.001     0.083   0.934        -0.001       0.001
     WkDay_1 |      -0.139      0.057    -2.457   0.014        -0.250      -0.028
     WkDay_2 |      -0.198      0.096    -2.065   0.039        -0.386      -0.010
     WkDay_3 |      -0.060      0.043    -1.411   0.159        -0.144       0.024
     WkDay_4 |      -0.034      0.036    -0.938   0.348        -0.105       0.037
     WkDay_5 |       0.007      0.043     0.155   0.877        -0.077       0.091
     WkDay_6 |      -0.032      0.016    -1.997   0.046        -0.064      -0.001
       _cons |       1.010      0.130     7.752   0.000         0.755       1.266
------------------------------------------------------------------------------
```

But ….

This much code is required to make a single graph

```
 summarize CTAS01_02
  local YScale1 = r(max) * 1.25
  summarize CTAS03
  local YScale2 = r(max) * 1.25
  summarize CTAS04_05
  local YScale3 = r(max) * 1.25
  display `YScale1' " "  `YScale2' " "  `YScale3'
  local YScale = max(`YScale1', `YScale2', `YScale3')
  * Create the Title(s)
  local Title1 "`NameZone' Zone"
  local Title2 "CTAS 1 & 2 combined, CTAS 3, and CTAS 4 & 5 combined"
  * Create the legend
  local Line1Text "CTAS 1 & 2 -- Actual (orange solid) and predicted (red dash)"
  local Line2Text "CTAS 3       -- Actual (blue solid) and predicted (purple dash)"
  local Line3Text "CTAS 4 & 5 -- Actual (green solid) and predicted (teal dash)"
twoway (line CTAS01_02 Date if (Date >= FirstPredict & Date <= LastCheck)
            , lcolor(orange) lwidth(thin) lpattern(solid)                    ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS01_02_Exp Date if (Date >= FirstPredict &                 ///
                                      Date <= LastPredict)                   ///
            , lcolor(red) lwidth(thin) lpattern(tight_dot)                   ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS01_02_Exp Date if (Date >= LastPredict &                  ///
                                      Date <= LastForecast)                  ///
            , lcolor(red) lwidth(thin) lpattern(dash)                        ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS03 Date if (Date >= FirstPredict &                        ///
                               Date <= LastCheck)                            ///
            , lcolor(navy) lwidth(thin) lpattern(solid)                      ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS03_Exp Date if (Date >= FirstPredict &                    ///
                                      Date <= LastPredict)                   ///
            , lcolor(purple) lwidth(thin) lpattern(tight_dot )               ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS03_Exp Date if (Date >= LastPredict &                     ///
                                      Date <= LastForecast)                  ///
            , lcolor(purple) lwidth(thin) lpattern(dash)                     ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS04_05 Date if (Date >= FirstPredict &                     ///
                               Date <= LastCheck)                            ///
            , lcolor(dkgreen) lwidth(thin) lpattern(solid)                   ///
              connect(direct) cmiss(n) )                                     ///
         (line CTAS04_05_Exp Date if (Date >= FirstPredict &                 ///
                                      Date <= LastPredict)                   ///
            , lcolor(teal) lwidth(thin) lpattern(tight_dot)                  ///
              connect(direct) cmiss(n))                                      ///
         (line CTAS04_05_Exp Date if (Date >= LastPredict &                  ///
                                      Date <= LastForecast)                  ///
            , lcolor(teal) lwidth(thin) lpattern(dash)                       ///
              connect(direct) cmiss(n) )                                     ///
, ytitle (Counts of CTAS Levels) ytitle(, size(small) )                      ///
  yscale(range(0 `YScale') )                                                 ///
  ylabel(#8, labsize(vsmall) format(%5.0f) ticks tlcolor(black)              ///
             tposition(outside) nogrid)                                      ///
  ymtick(##10, ticks tlcolor(black) tposition(outside) nogrid)               ///
  xtitle(Calendar Date) xtitle(, size(small))                                ///
  xlabel(#5, labsize(vsmall) labcolor(black) ticks  tlcolor(black)           ///
           tposition(outside) grid)                                          ///
  xmtick(#60, ticks tlcolor(black) tposition(outside) nogrid)                ///
  title("`Title1'" "`Title2'"                                                ///
        , size(small))                                                       ///
  note( "`Line1Text'" "`Line2Text'"  "`Line3Text'"                           ///
        , size(vsmall))                                                      ///
  legend(off)
```



and it takes a lot of tweaking …

and ….

This much code is required for data cleaning

```
gen str FacilityCode = substr(Facility, 1, 3), after (Facility)
replace FacilityCode = FacilityCode + "-" + SiteCB if(FacilityCode == "CBH")
label variable FacilityCode "Facility Code (Mnemonic)"
tabulate FacilityCode

* The encode command is convenient
* It will create numerical assignments in file order.
*    encode FacilityCode, gen (FacilityNum)
*       IT ADDS VALUE LABELS AUTOMATICALLY
*    tabulate FacilityNum, nolabel
*    order FacilityNum, after(FacilityCode)
* However, the following will provide alphabetical or custom orders
*   Herein the regional facilities are listed first
*   In alphabetical order (by full name), with 4 sites for CBR
*
* Numerical codes are required because some procedures (eg collapse)
* are restricted to numeric values.

gen FacilityNum = 0, after(FacilityCode)
label variable FacilityNum "Facility Coded by Number (see labels)"
replace FacilityNum = 1  if ( FacilityCode == "ARH" )
replace FacilityNum = 2  if ( FacilityCode == "CBH-CBG" )
replace FacilityNum = 3  if ( FacilityCode == "CBH-CBN" )
replace FacilityNum = 4  if ( FacilityCode == "CBH-CBR" )
replace FacilityNum = 5  if ( FacilityCode == "CBH-CBW" )
replace FacilityNum = 6  if ( FacilityCode == "CRH" )
replace FacilityNum = 7  if ( FacilityCode == "CCH" )
replace FacilityNum = 8  if ( FacilityCode == "SMR" )
replace FacilityNum = 9  if ( FacilityCode == "SSR" )
replace FacilityNum = 10 if ( FacilityCode == "VRH" )
replace FacilityNum = 11 if ( FacilityCode == "YRH" )
replace FacilityNum = 12 if ( FacilityCode == "ACH" )
replace FacilityNum = 13 if ( FacilityCode == "ASH" )
replace FacilityNum = 14 if ( FacilityCode == "BMH" )
replace FacilityNum = 15 if ( FacilityCode == "DGH" )
replace FacilityNum = 16 if ( FacilityCode == "EMH" )
replace FacilityNum = 17 if ( FacilityCode == "FMH" )
replace FacilityNum = 18 if ( FacilityCode == "GMH" )
replace FacilityNum = 19 if ( FacilityCode == "ICH" )
replace FacilityNum = 20 if ( FacilityCode == "LFM" )
replace FacilityNum = 21 if ( FacilityCode == "NCM" )
replace FacilityNum = 22 if ( FacilityCode == "QGH" )
replace FacilityNum = 23 if ( FacilityCode == "RWH" )
replace FacilityNum = 24 if ( FacilityCode == "SAC" )
replace FacilityNum = 25 if ( FacilityCode == "SCC" )
replace FacilityNum = 26 if ( FacilityCode == "SHH" )
replace FacilityNum = 27 if ( FacilityCode == "SMH" )
replace FacilityNum = 28 if ( FacilityCode == "SMM" )
replace FacilityNum = 29 if ( FacilityCode == "SRH" )
replace FacilityNum = 30 if ( FacilityCode == "VCM" )
tabulate FacilityNum
label define FacilityNum   1 "ARH: Pictou County Health Authority", modify
label define FacilityNum   2 "CBH-CBG: Cape Breton Healthcare Complex, Glace Bay", modify
label define FacilityNum   3 "CBH-CBN: Cape Breton Healthcare Complex, Northside", modify
label define FacilityNum   4 "CBH-CBR: Cape Breton Healthcare Complex, CBR", modify
label define FacilityNum   5 "CBH-CBW: Cape Breton Healthcare Complex, New Waterford", modify
label define FacilityNum   6 "CRH: Colch East Hants Health Center", modify
label define FacilityNum   7 "CCH: Cumberland Reg Healthcare Ctr", modify
*/
tabulate Triage if ( Triage_Date == . & DepartureStatus ==                  ///
                        "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
Total |          1      100.00
*/

count if ( Triage < . & Triage_Date < . & DepartureStatus ==                ///
                        "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
replace CommentsTime = "Triage: Flag Only - left without being triaged, " +  ///
                       "but has triage time and value"                      ///
                       if ( Triage < . & Triage_Date < . & DepartureStatus   ///
                            == "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
count if ( Triage == . & Triage_Date < . & DepartureStatus ==               ///
                        "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
replace CommentsTime = "Triage: Flag Only - left without being triaged, " +  ///
                       "but has triage time but no value"                   ///
                       if ( Triage == . & Triage_Date < . & DepartureStatus  ///
                            == "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
count if ( Triage < . & Triage_Date == . & DepartureStatus ==               ///
                        "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
replace CommentsTime = "Triage: Flag Only - left without being triaged, " +  ///
                       "but has triage value but no time"                   ///
                       if ( Triage < . & Triage_Date == . & DepartureStatus  ///
                            == "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
save "MT_FY2015-18_Cleaned", replace


save "MT_FY2015-18_Cleaned", replace
```

```
label define FacilityNum   8 "SMR: St. Martha's Regional Hospital", modify
label define FacilityNum   9 "SSR: South Shore Regional", modify
label define FacilityNum  10 "VRH: Valley Regional Hospital", modify
label define FacilityNum  11 "YRH: Yarmouth Regional Hospital", modify
label define FacilityNum  12 "ACH: Annapolis Community Health Ctr", modify
label define FacilityNum  13 "ASH: All Saints' Springhill Hosp", modify
label define FacilityNum  14 "BMH: Buchanan Memorial Hospital", modify
label define FacilityNum  15 "DGH: Digby General Hospital", modify
label define FacilityNum  16 "EMH: Eastern Memorial Hospital", modify
label define FacilityNum  17 "FMH: Fishermen's Memorial Hospital", modify
label define FacilityNum  18 "GMH: Guysborough Memorial Hospital", modify
label define FacilityNum  19 "ICH: Inverness Consolidated Mem", modify
label define FacilityNum  20 "LFM:  Lillian Fraser Memorial Hosp", modify
label define FacilityNum  21 "NCM: North Cumberland Memorial Hosp", modify
label define FacilityNum  22 "QGH: Queens General Hospital", modify
label define FacilityNum  23 "RWH: Roseway Hospital", modify
label define FacilityNum  24 "SAC: St Anne's Community Centre", modify
label define FacilityNum  25 "SCC: South Cumberland Comm Care Ctr", modify
label define FacilityNum  26 "SHH: Sacred Heart Hospital", modify
label define FacilityNum  27 "SMH: Soldiers' Memorial Hospital", modify
label define FacilityNum  28 "SMM: St. Mary's Memorial Hospital", modify
label define FacilityNum  29 "SRH: Strait-Richmond Hospital", modify
label define FacilityNum  30 "VCM: Victoria County Memorial Hosp", modify
label values FacilityNum FacilityNum
tabulate FacilityNum FacilityCode
tabulate FacilityNum
tabulate FacilityNum, nolabel

gen SiteCBNum = ., after(SiteCB)
replace SiteCBNum = 1 if (SiteCB == "CBR")
replace SiteCBNum = 2 if (SiteCB == "CBG")
replace SiteCBNum = 3 if (SiteCB == "CBN")
replace SiteCBNum = 4 if (SiteCB == "CBW")
label variable SiteCBNum "CB Site: CBR (1), GB=CBG (2), NS=CBN (3), NW=CBW (4)"
label define SiteCB 1 "CBR: Cape Breton Regional ED"
label define SiteCB 2 "CBG: Glace Bay ED", add
label define SiteCB 3 "CBR: Northside ED", add
label define SiteCB 4 "CBR: New Waterford ED", add
label values SiteCBNum SiteCB
tabulate SiteCB SiteCBNum

count if (Zone == . )
* 3,374
display _N - r(N)
* 1170206
tabulate Zone
* missing 3374
tabulate FacilityNum if ( Zone == .), nolabel
tabulate Zone if ( FacilityNum == 24 )
* Missing on all
replace Zone = 3 if ( FacilityNum == 24 )
* Done

count if (HCN == . )
display _N - r(N)
* missing on 196

count if (Sex == "" )
tabulate Sex
* No missing values -- some
gen SexNum = ., after (Sex)
replace SexNum = -1
replace SexNum = 0 if (Sex == "F" | Sex == "f" )
replace SexNum = 1 if (Sex == "M" | Sex == "m" )
label variable SexNum "Sex: 1 = Male, 0 = Female)
tabulate SexNum Sex

count if ( Age == . )
* No missing values

count if ( Residence == "" )
* Missing on 4995
gen int ResidenceCode = real(substr(Residence, 1, 2)), after(Residence)
label variable ResidenceCode "Place of Residence coded by number (see labels)"
gen int ResidenceSimple = ., after (ResidenceCode)
replace ResidenceSimple = 1 if ( ResidenceCode < 80)
replace ResidenceSimple = 2 if ( ResidenceCode > 79 & ResidenceCode < 97 )
replace ResidenceSimple = 3 if ( ResidenceCode == 97 )
replace ResidenceSimple = 4 if ( ResidenceCode == 98 )
replace ResidenceSimple = 5 if ( ResidenceCode == 99 )
replace ResidenceSimple = 5 if ( ResidenceCode == . )
label value ResidenceSimple "Simplified Place of Residence: NS, CAN, US, Other,
Uknown/Missing"
label define Residence 1 "Nova Scotia"
label define Residence 2 "Canada other than Nova Scotia", modify
label define Residence 3 "USA", modify
label define Residence 4 "Other Countries", modify
label define Residence 5 "Unknown or Missing", modify
tabulate Residence ResidenceSimple
count if ( ResidenceSimple == . )
* missing on 4,995
```

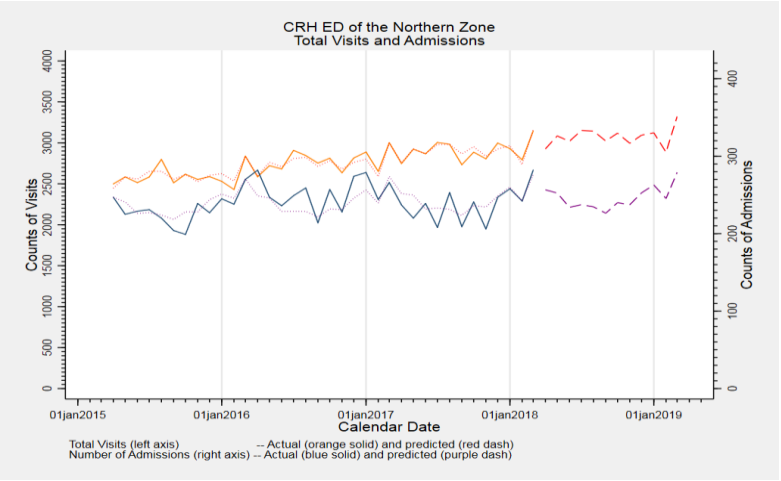
```
* Not done here down... more like a "template" depending on what is needed.

count if ( FamilyDoctor == "" )
* Missing on 29776
count if ( FamilyDoctorCode == "" )
* Missing on only 23788
* Fix later when needed.

display _N
tabulate Insurance
* missing 1
tabulate InsuranceCode
* missing 2
tabulate Insurance if ( InsuranceCode == "" )
tabulate Insurance InsuranceCode
replace InsuranceCode = "DOH" if ( InsuranceCode == "" & Insurance != "" )

list ResidenceSimple if ( InsuranceCode == "" | Insurance == "" )
replace InsuranceCode = "DOH" if ( InsuranceCode == "" & Insurance == "" )

count if ( Scheduled == "" )
tabulate Scheduled
gen ScheduledNum = 0, after (Scheduled)
label variable ScheduledNum "Scheduled: Yes (1) or No (0), No assumed for missing"
replace ScheduledNum = 1 if ( Scheduled == "Y" | Scheduled == "y" )
tabulate ScheduledNum
tabulate ScheduledNum Scheduled

tabulate PresentingComplaint
* All have values, but 308,870 are "NULL"
tabulate FacilityNum if ( PresentingComplaint == "NULL" )
* Spread across all facilities.

tabulate AdmissionSource
* All have values, but 998,956 are "NULL"
* Another 46,811 are "UNKNOWN"

save "MT_FY2015-18_V2", replace
* use "MT_FY2015-18_V2", clear

* Basic fixes for times
* Times are the core to visits ... correct the obvious errors, and flag the rest

count if ( Arrival_Date == . )
* Missing on 127,186
count if ( Triage_Date == . )
* Missing on 2,565
count if ( Regist_Date == . )
* Missing on 0
count if ( AssRoom_Date == . )
* Missing on 709,492
count if ( PhysAssess_Date == . )
* Missing on 409,053
count if ( OBS_In_Date == . )
* "Missing" on 1,156,862
count if ( OBS_Out_Date == . )
* "Missing" on 1,156,894
count if ( OBS_DisChg_Date == . )
* "Missing" on 1,161,004
count if ( CDU_In_Date == . )
* "Missing" on 1,156,862
count if ( CDU_Out_Date == . )
* "Missing" on 127,186
count if ( ADT_Decision_Date == . )
* "Missing" on 1,136,789
count if ( ADT_In_Date == . )
* "Missing" on 1,109,334
count if ( ADT_Out_Date == . )
* "Missing" on 1,095,125
count if ( ADT_OutBed_Date == . )
* "Missing" on 1,096,556
count if ( ADT_DisChg_Date == . )
* "Missing" on 1,168,079
count if ( Depart_Date == . )
* Missing on 5,765
count if ( LeftER_Date == . )
* Missing on 4,921

gen str CommentsTime = "", after(LeftER_Date)
* Comments have the form:
* TimeUnit: Flag Only - error
* TimeUnit: Fixed as ... - error

* Registration is the reference point.
* Registation time should equal the triage and arrival dates -- more or less.
count if ( Triage_Date == . )
* 2565
tabulate DepartureStatus if ( Triage_Date < . )
tabulate Triage if ( Triage_Date < . & DepartureStatus ==                   ///
```

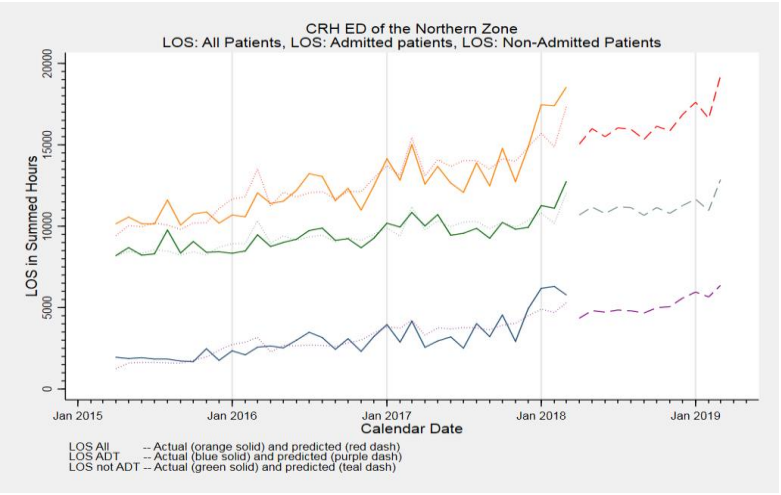
```
tabulate Triage if ( Triage_Date == . )
Clearly "triaged": Set date to Regist_Date
count if ( Triage_Date == . & Triage < . )
replace CommentsTime = "Triage value but no date: "+                        ///
                       "Replaced Triage date with Registration date"        ///
                       if ( Triage_Date == . & Triage < . )
replace Triage_Date = Regist_Date if ( Triage_Date == . & Triage < . )
save "MT_FY2015-18_Cleaned", replace

* Can trust admission dates -- they are "important"
count if ( Triage_Date == . & Triage ==  . & ADT_Decision_Date < . )
* 19
* These have been admitted -- hence triaged (implicitly, at least)
replace CommentsTime = "Triage: Fixed Admitted but no triage date or value: "+     ///
                       "Replaced Triage date with Registration date"        ///
                     if ( Triage_Date == . & Triage < . & ADT_Decision_Date < .)
replace Triage_Date = Regist_Date if ( Triage_Date == . & Triage ==  . &     ///
                                       ADT_Decision_Date < . )

count if ( Triage_Date == . & Triage ==  . & Depart_Date < . )
* 65

tabulate DepartureStatus if ( Triage_Date == . & Triage ==  . & Depart_Date < . )
* same as ...
tabulate DepartureStatus if ( Triage_Date == . & Triage ==  . )
count if ( Triage_Date == . & Triage ==  . & DepartureStatus != "" &         ///
          DepartureStatus != "NCR NT - PT LT BEFORE BEING TRIAGED(OP)")
replace CommentsTime = "Discharged (not LWBS) but no triage date or "+      ///
                       "value: Replaced Triage date with Registration date"  ///
                       if ( Triage_Date == . & Triage ==  .                  ///
                          & DepartureStatus != "" &                          ///
                     DepartureStatus != "NCR NT - PT LT BEFORE BEING TRIAGED(OP)")
replace Triage_Date = Regist_Date                                           ///
                      if ( Triage_Date == . & Triage ==  .                  ///
                          & DepartureStatus != "" &                         ///
                     DepartureStatus != "NCR NT - PT LT BEFORE BEING TRIAGED(OP)")

* Patients who "should have been" triaged (at least implicitly)
count if ( Triage_Date == . & Triage ==  . & AssRoom_Date < .)
* 315
count if ( Triage_Date == . & Triage ==  . & PhysAssess_Date < .)
* 0
count if ( Triage_Date == . & Triage ==  . & CDU_In_Date < .)
* 6
count if ( Triage_Date == . & Triage ==  . & OBS_In_Date < .)
* 123
replace CommentsTime = "On treatment path but no triage date or value: "+   ///
                       "Replaced Triage date with Registration date"        ///
                       if ( Triage_Date == . & Triage ==  . &               ///
                            (AssRoom_Date < . | PhysAssess_Date < . |        ///
                             CDU_In_Date < .  | OBS_In_Date < .) )
replace Triage_Date = Regist_Date                                           ///
                      if ( Triage_Date == . & Triage ==  . &                ///
                           (AssRoom_Date < . | PhysAssess_Date < . |        ///
                            CDU_In_Date < .  | OBS_In_Date < .) )
tabulate DepartureStatus if ( Triage < . )


tabulate PresentingComplaint if ( Triage_Date == . & Triage ==  .           ///
                                  & PresentingComplaint != "NULL" )
tabulate DepartureStatus if ( Triage_Date == . & Triage ==  .               ///
        & DepartureStatus == "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )
replace Triage_Date = Regist_Date if ( Triage_Date == . & Triage ==  .      ///
        & DepartureStatus != "NCR NT - PT LT BEFORE BEING TRIAGED(OP)" )

* Numeric versions of variables that are "generic" and of likely us in any
* analysis.
* These include Sex, Admitted (yes=1), CDU patient (yes=1) and
* Observation patient (yes=1)
* Age is already numeric. Age groups depend on the analysis so are not included
* here
* Residence is coded as NS, vs Canada, vs USA, vs Other. Analyses below the
* level of NS will depend on the institution(s) used.
```

… it goes on and on and on …

What Have We Done?

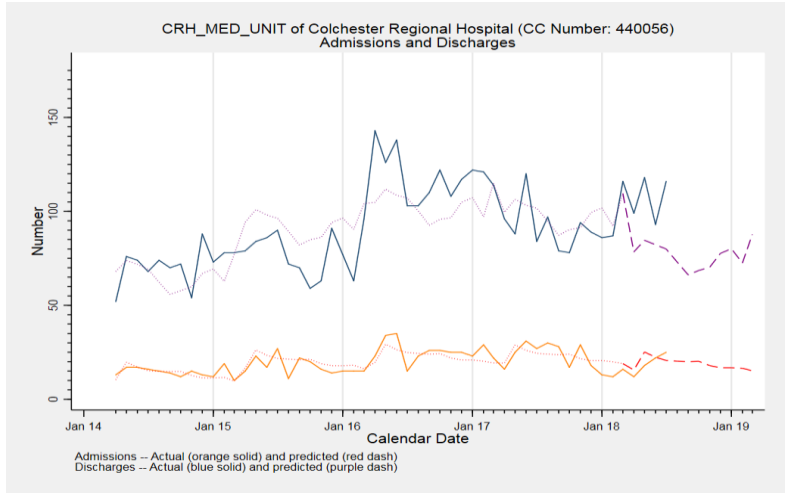
ED Visits, Admissions, LOS

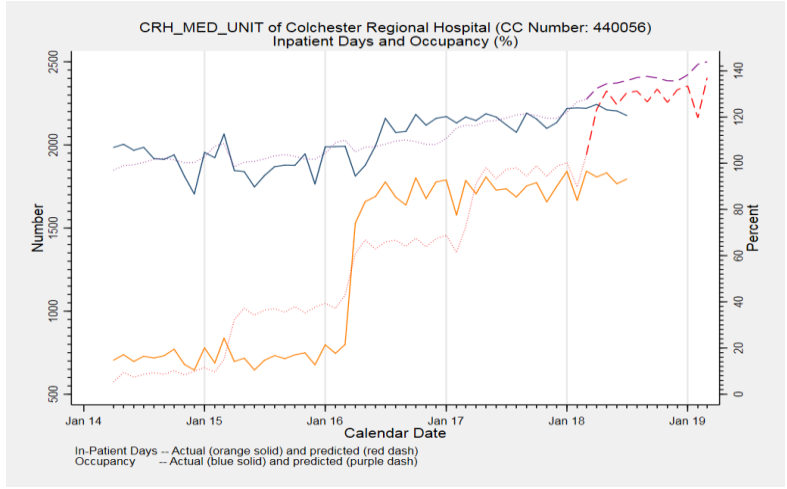




"all"

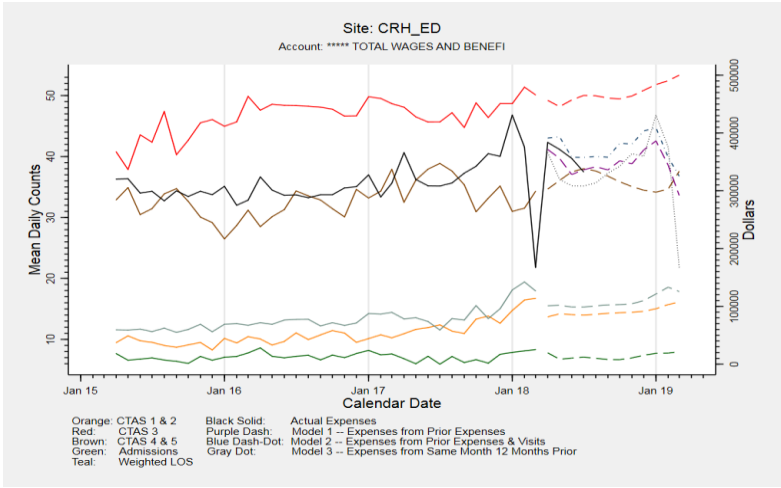
In-Patient, Admissions, Transfers, Discharges, LOS

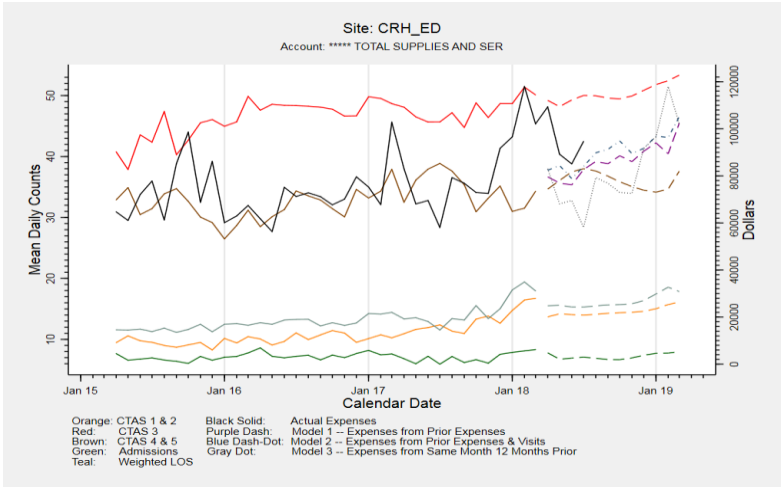




"all"

Expenses by Unit and Account

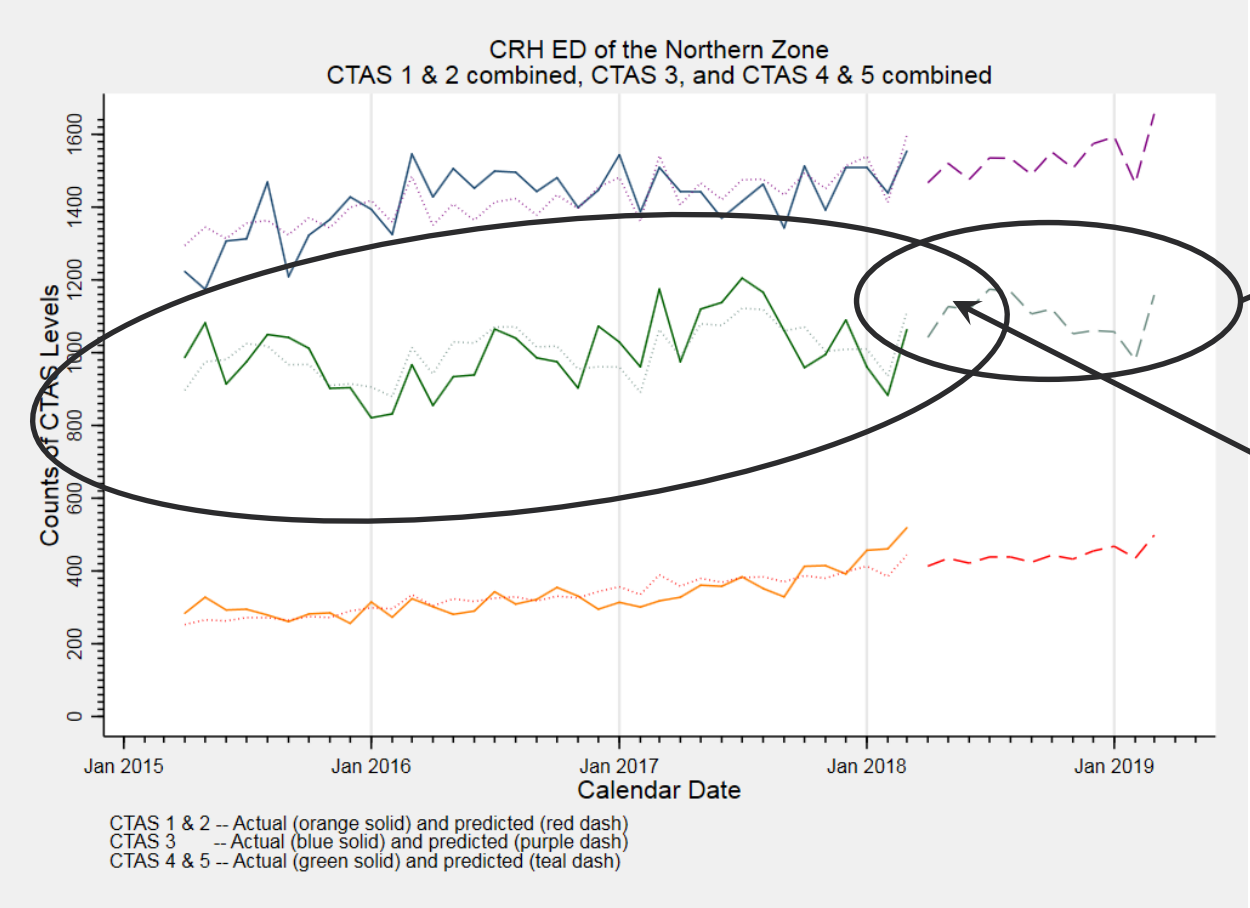




"pilot"

Forecasting

Using an Understanding of the Past to Predict the Future



Forecasting

- A basic extension of the model to future dates
- All IV (predictors) of the model must also exist in the future.

Testing

- Models should always be tested on data that was NOT used to build the model.
- Fortunately, data is always coming in.

Forecasting

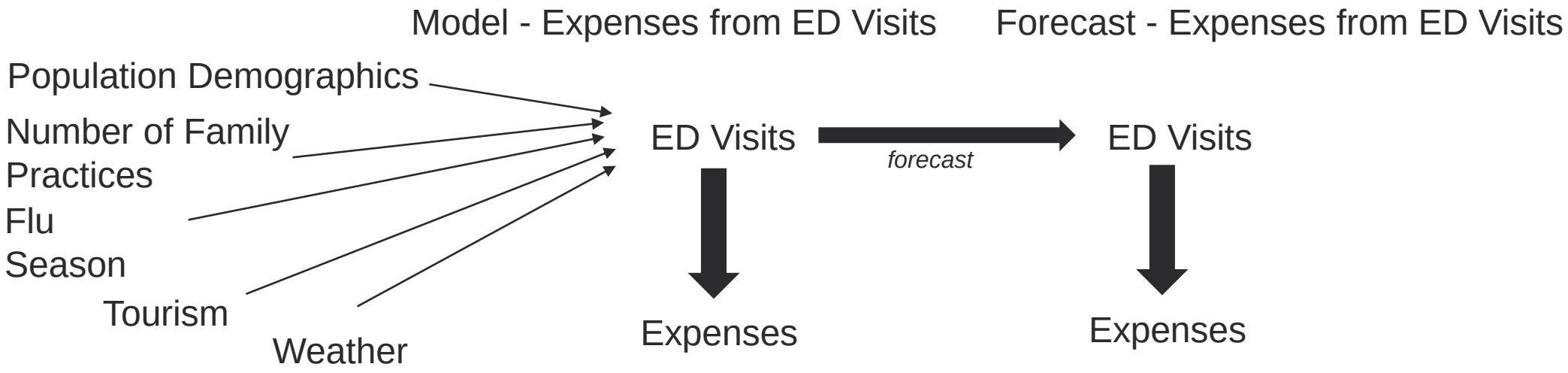
The quality (utility) of the forecast depends on the quality of the model,

But

Forecasting is not the same as modelling.

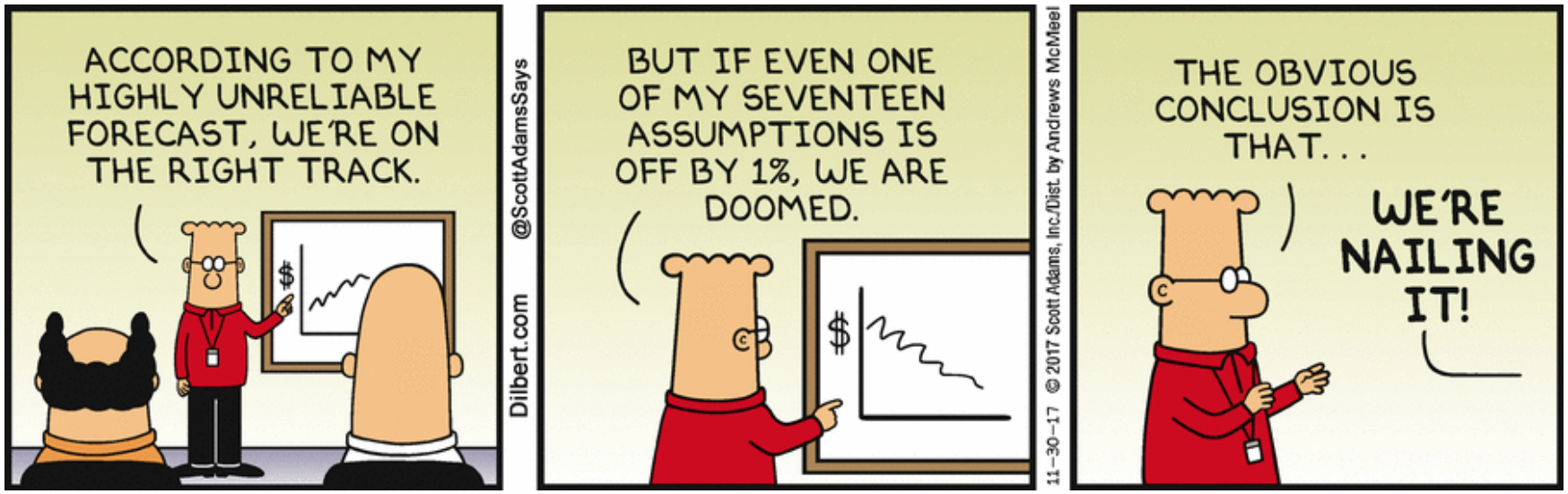
All the variables used in the model must exist in the future.

- Must be KNOWN.
- Must be *predictable*.



Modelling and Forecasting

Assessing the Analysis



- Can you describe the sources of data in the analysis?
- Are you sure the data (sample data) is representative of the population?
- Are their outliers in the data distribution? How did they affect the results?
- What assumptions are behind your analysis?
- Are there conditions that would make your assumptions invalid?

Modelling and Forecasting Next Steps

Lots to be Done

- New variables / better variables
- New models / better models
- Application to other services
- Partnership with other teams

Applied Analytics | Utilization Forecasting - Structural Equation Modelling

Business Problem
Lack of forecasts of health system utilization to support planning

Solution
Utilization SEM analysis based on multiple interacting regression models

Business Value
Enhanced forecasts leading to operational improvements and improved patient experiences

Data

Data: Inpatient and ED administrative data for key utilization metrics, demographics, census, policy changes, vital stats, etc.

Applied Analytics

Data Analysis

Information Exploration

Data Analysis

Modelling and Validation

Structural Equation Modelling

Action

- Dissemination to leaders
- Dynamic Reporting
- Automation

Data Requirement

Record level and aggregate data from multiple systems and sources, complex ELT processes to clean and normalize data

Modeling

Examination of multiple models to produce analysis

Path Analysis

Confirmatory Factor Analysis

Structural Equation Modeling

Fuzzy Analytics Hierarchy Process

(abstracting a system goal for optimization purpose)

Improved Forecasting

Potential Short-Term Forecasting

Explaining differences between areas

- Cross area spill-over effects
- Demand fluctuation
- Population emergency sensitivity
- Boarding process efficiency
- Bed substitutability

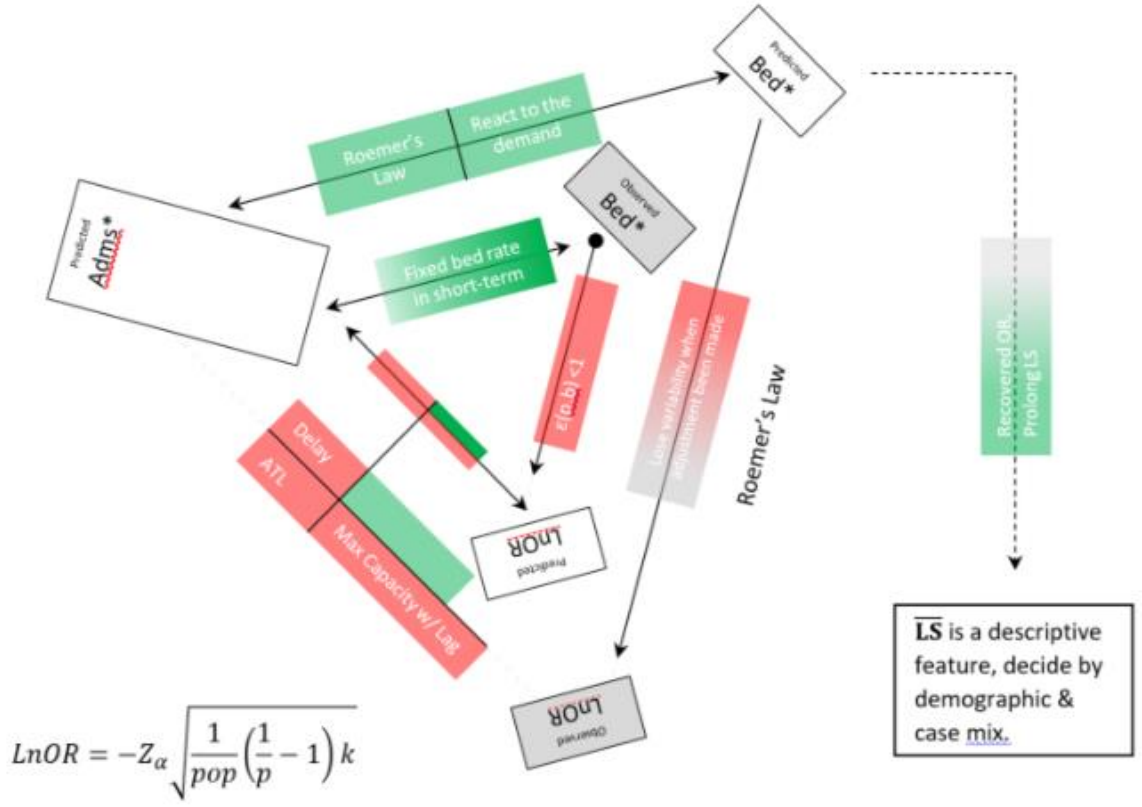
Potential Long Term Forecasting

Identify effects both temporal and spatial wise

Bed pooling strategy

System-wide optimization

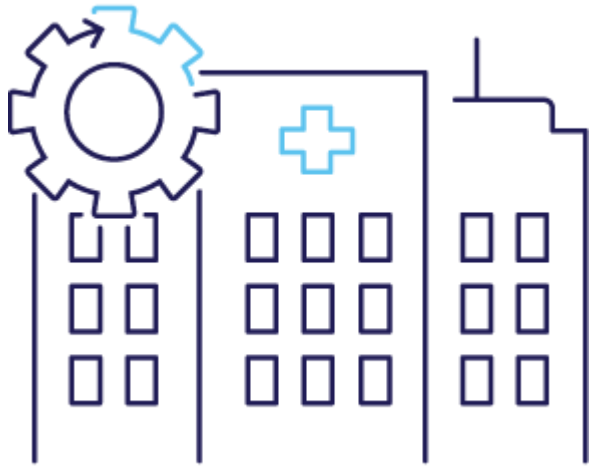
System Utilization Forecasting using Structural Equation Modeling

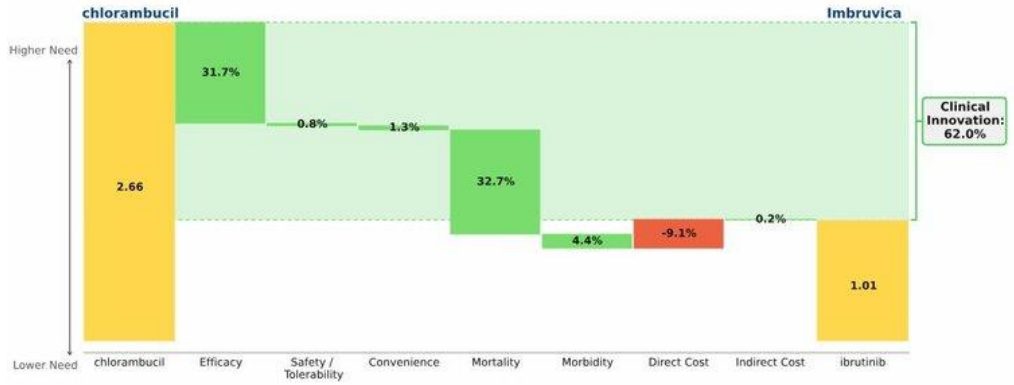


- The purpose of this project is to develop a predictive model to investigate and quantify the **interactions** among key utilization indicators both in short (inner) and long (outer) term.
- Focused on exploring the demand and supply dynamics behind the everyday individual activities using structural equation modeling

System Utilization Forecasting using Structural Equation Modeling

- This project will introduce a different methodology with additional explanatory factors.
- The model will be used to identify key attributes for explaining historical congestion events, identifying sub-optimal areas and conducting optimization and “what-if” scenario simulation.





Applied Analytics | Population Based Utilization Projections

Business Problem
Lack of forecasts of health system utilization to re-development support planning

Solution
Population based projections based on QEII Re-development model

Business Value
Population based projections to support robust decision making

Data

Data Sources

Record level administrative data across key functional areas

Applied Analytics

Data Analysis

Information Exploration

Data Analysis

Data Manipulation and Validation

Programming and Descriptive Statistics

Action

Dissemination to leaders

Dynamic Reporting

Adjustment of results

Data Requirement

Record level data identifying clinical details, demographics and geography

Modeling

Population models to produce forecasts

Data management and programming problem with some modeling of population

Improved Forecasting

Potential Short-Term Planning

Data to support functional and master planning for re-development

Potential Long Term Planning

Greater understanding of what we can expect if we deliver services in current models of care

Our Approach

Multi-step analysis of historical and projected data as inputs

This approach is focused on providing as valid as possible a sense of the future need within the catchment area, through an iterative (but, consistent algorithmic) process to meet the project and other planning needs.

1

POPULATION PROJECTIONS

Utilized NS Dept. of Finance projections to 2028 by sex and 5 year age groupings for NS at the county level, projections extended to 2035.

2

UTILIZATION PROJECTIONS

To project (forecast) the utilization, there were two main processes. “Baseline projections” and “adjusted projections”. The main problem with both is not the process, rather the number of variables to be considered.

3

BASELINE PROJECTIONS

The first step computed the percentage of the population that used each service. These computations were conducted within each age group, sex and catchment area separately. The count was then projected into the future within each grouping.

4

ADJUSTED PROJECTIONS

The adjusted projections are not yet started. The adjusted projections take the baseline projections and then adjust based on in year and historical trends as well is adjustment factors around clinical or patient flow efficiency.

5

ITERATIVE ANALYSIS

This analysis is designed to be iterative, we will deliver the baseline projections and adjust the analysis to meet the project needs as necessary.

Population Projections

Statistics Canada:

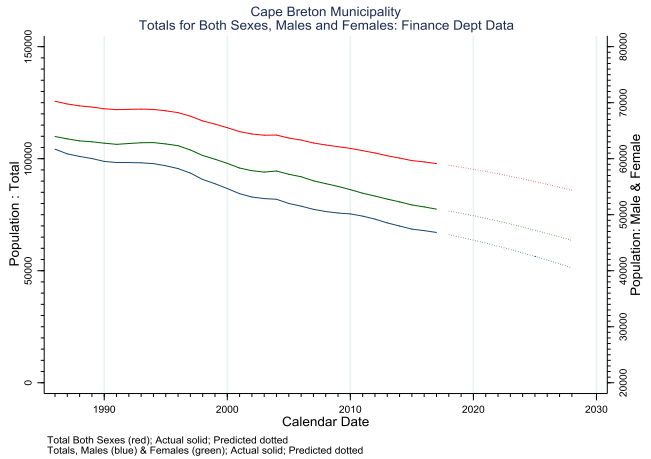
- NS historical population data from 1986 to 2017
- by sex and age group

NS Department of Finance and Treasury Board:

- converted to the county level
- projections from 2018 to 2028

NSHA Performance and Analytics

- converted to catchment areas
  - CBRM
  - Rest of Cape Breton
  - Rest of Nova Scotia
  - Canada
  - Other
- extended to 2035 within each Sex, Age Group, and Catchment area



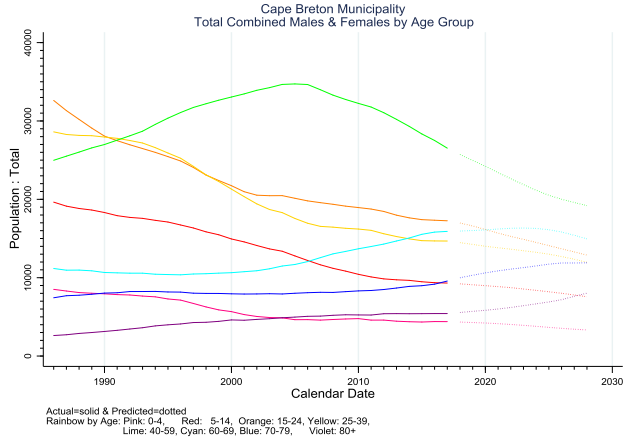
From Population to Utilization Projections

Age groups were collapsed
- adjacent levels combined
- retained the Statistics Canada boundaries
- clinically relevant age groups

Final Population Categories
- 5 Catchment Areas by 8 Age Groups by 2 Sexes
- 80 combinations

Converted Population Level to Change in Population Relative to 2017
- percentage of the population in 2017

Forecasted utilization was computed as:
- %ΔUtilization = %Δ population * utilization in 2017
  - %Δ is percentage change
  - utilization is "visits" or "admissions" or "discharges" etc.
  - population is each Sex by Age Group by Catchment Area



Example: Forecasting Utilization in the CB ED

Number of Patient Visits

Population in 2017, and Population Forecasts

- grouped by:
  - catchment areas
  - age groups
  - sex

Emergency Patient Data from Meditech

- data cleaning
- grouped by:
  - catchment areas
  - age groups
  - sex

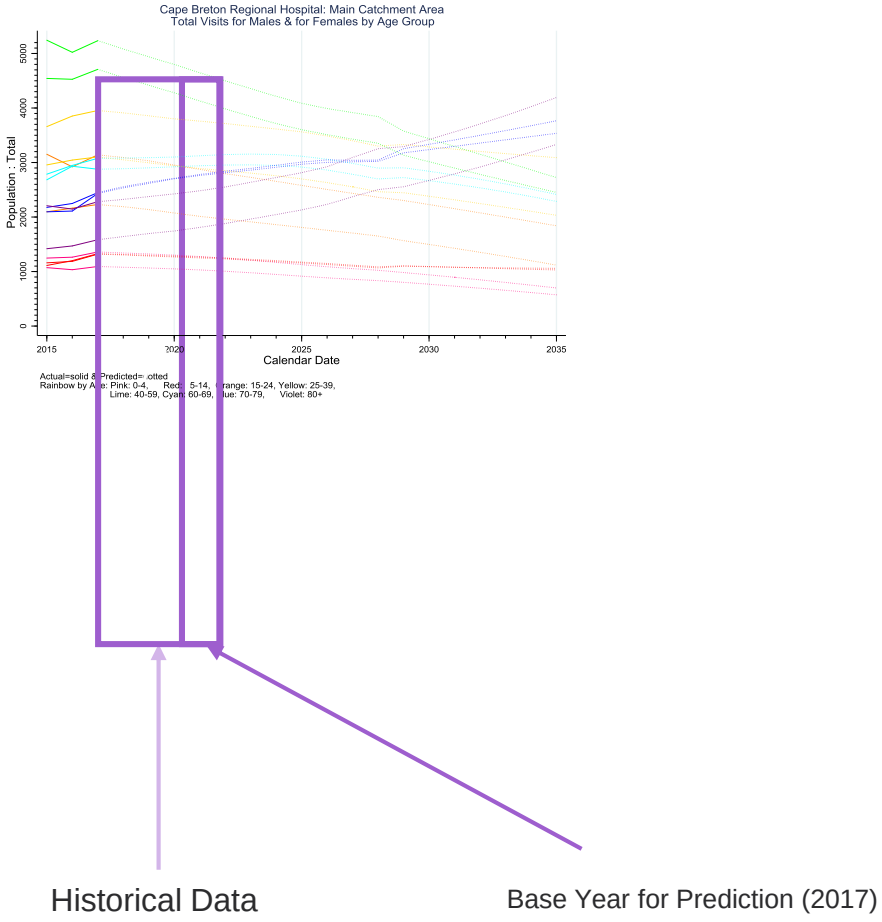
Visits in 2017

- convert to proportion of population
- create forecast

Forecasted Visits in FY2018 – 2035:

$Visits_{Future} = Population_{Future} * Visits_{2017} / Population_{2017}$

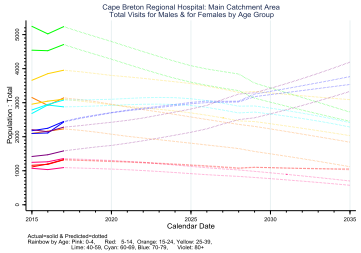
- With each age and sex group



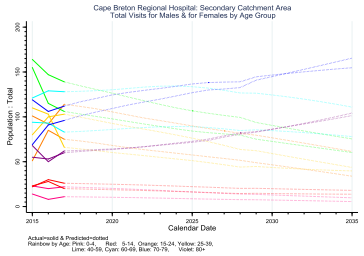
Example: Forecasting Utilization in the CB ED

Repeat for Catchment Areas 2 & 3

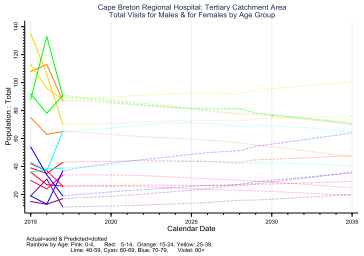
Main Catchment Area
Cape Breton Regional Municipality



Secondary Catchment Area
Victoria, Inverness, Richmond Counties



Tertiary Catchment Area
Rest of NS



Collapse

over catchment area, age group, and sex

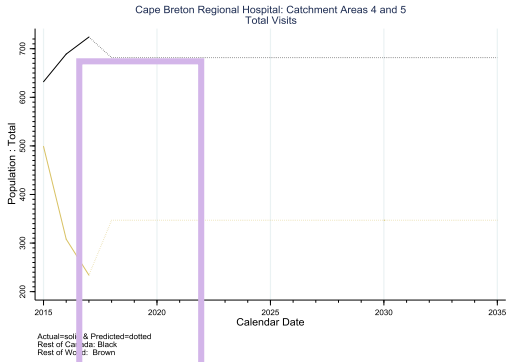
Example: Forecasting Utilization in the CB ED
Catchment Areas 4 (Canada) and 5 (World)

Catchment Areas 4 & 5

1. ~~Cape Breton Regional Municipality~~
2. ~~Victoria, Inverness and Richmond Counties~~
3. ~~Nova Scotia excluding Cape Breton~~
4. Canada excluding NS
5. the world excluding Canada

Get Utilization

1. collapse over sex and age groups
2. assess average number of visits prior to 2018
3. use that average as the forecast

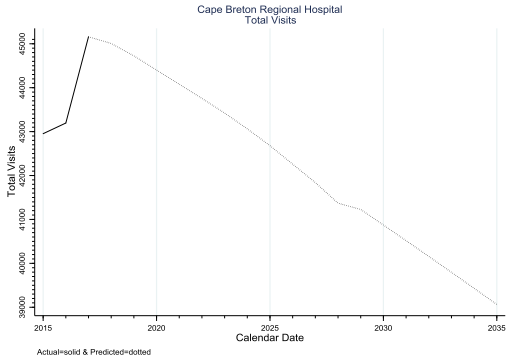


Average of Historical Data

Example: Forecasting Utilization in the CB ED
Final Forecast for Visits

Have Forecasts for Catchment Areas 1, 2, 3, 4 & 5

1. Cape Breton Regional Municipality
2. Victoria, Inverness and Richmond Counties
3. Nova Scotia excluding Cape Breton
4. Canada excluding NS
5. the world excluding Canada



Collapse

collapse over all catchment areas

Repeat for other Facilities

- New Waterford Consolidated Hospital,
- Northside General Hospital,
- Glace Bay Hospital

Repeat for other Utilization Metrics

- Visits by CTAS, Admissions, LOS, Required Beds

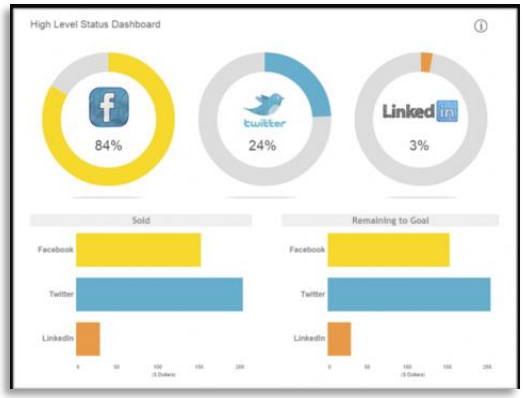
Adjustments

Forecasts were packaged as Excel and Tableau files

- Presented to relevant redevelopment teams

Various meetings (& emails) to:

- explain the process
  - particularly the “roll-up” of services
- assess the “validity” of the results
  - “face validity”
  - various checks
    - most teams could provide their own analysis of visits
    - current bed requirements vs forecasted



Adjustments

- resolving discrepancies
- combining or dropping particular services
  - often related to observations of data prior to 2017
- splitting services into finer categories
- use of a different base year

Re-analysis

- not usually the entire process
  - there were built-in “break points”

More meetings

General Comments on the Population Based Projections Process

A lot of data manipulation “programming”, not a lot of statistics

Could not be done in Excel, was done in Stata

- could be done in SAS, SPSS, R, C/C++, Python, etc.

Requires automation for efficiency and scale in the interim relies on re-usable code

- automation is required
  - the volume of work
  - generalizations across units/services
- process is iterative at all stages
  - data cleaning
  - roll-ups
  - need to re-run analyses several times

Applied Analytics | Predicting Patient Flow

Business Problem
Develop model to predict arrival rate, type and patient path and package information into actionable insight for leaders

Solution
Machine Learning Proof of Concept to apply power and scale to complex problem

Business Value
Translating predictive analytics into a change in practice at site level

Data

Data: Inpatient and ED administrative data at an encounter level for multiple years (millions of records potentially)

Applied Analytics

Data Analysis

Information Exploration

Data Analysis

Modelling and Validation

Machine Learning Models

Action

- Dissemination to leaders
- Dynamic Reporting
- Automation/Scale

Data Requirement

Record level ED and Inpatient data for pilot sites, complex ELT processes to prepare and anonymize data

Modeling

Examination of multiple models to produce analysis

Classification

Prediction

Clustering

Understanding flow of patients and patient level bottleneck in a dynamic and scalable model

Improved Forecasting

Potential Short-Term Forecasting

Predicting LOS and Outcomes at patient level

Potential Long Term Forecasting

Identify clusters of patients with common characteristics and planning appropriate interventions

nova scotia health authority

Predicting Patient Flow

Proof of Concept and EXTRA Project

- Predictive statistical models have been shown to predict patient flow with fairly high accuracy

- There have also been specific examples, including in the Canadian context, of how applied analytics can improve understanding of an impact of patient flow and decision-making.

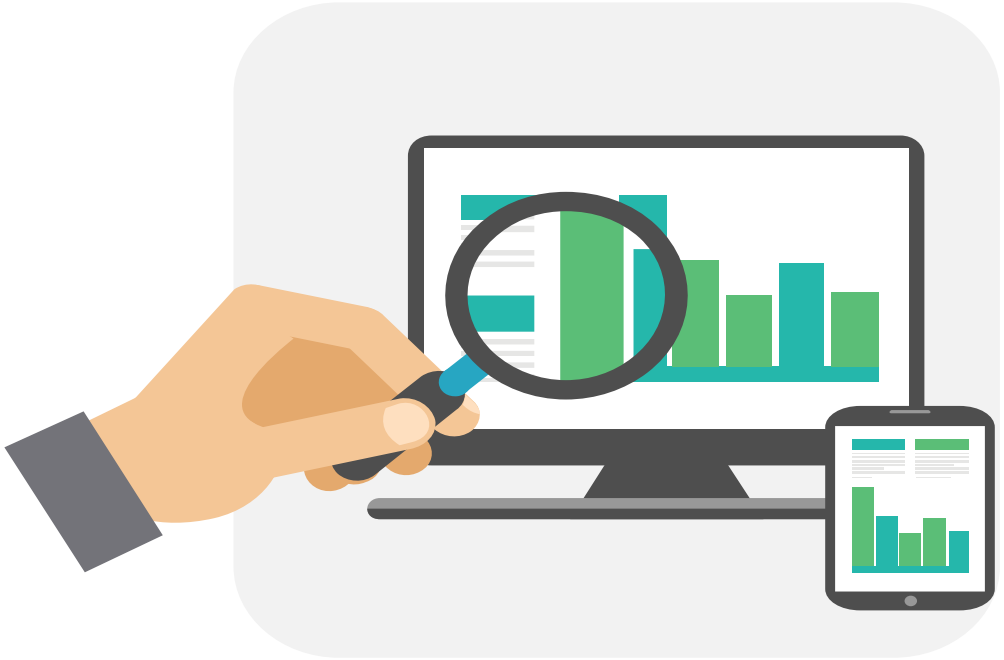
  - Humber River Hospital Command Center – this is describes as a ‘data-driven mission control’: https://www.hrhfoundation.ca/commandcentre/
  - Institute for Clinical Evaluative Sciences: https://www.ices.on.ca/Publications/Journal-Articles/2012/January/Evaluating-the-effect-of-clinical-decision-units-on-patient-flow-in-seven-Canadian-emergency, https://www.ices.on.ca/Publications/Journal-Articles/2017/July/Emergency-department-flow-measures-for-adult-and-pediatric-patients-in-British-Columbia-and-Ontario
  - Presentation at the most recent HIMSS international conference titled ‘Deriving value from patient flow analytics’: https://www.himssconference.org/session/deriving-value-patient-flow-analytics

Applicability and Feasibility

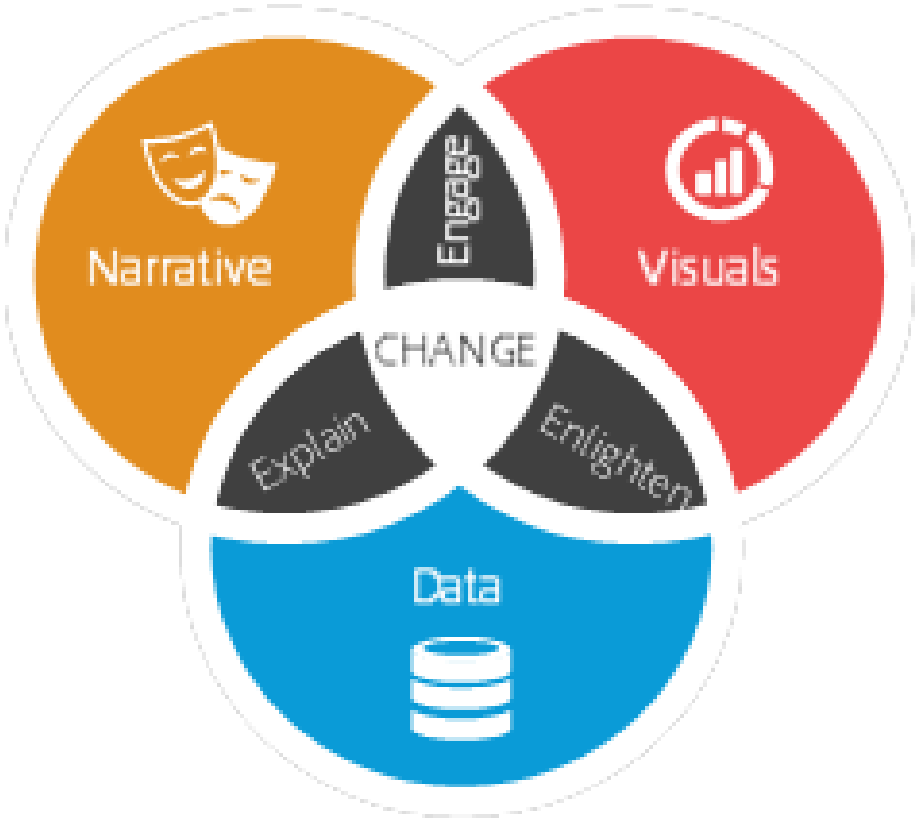
- This improvement project is one of the first at NSHA.
- Prior to now, we have had limited analytical capacity to take on this time of project.
- The skill sets of our team members in terms of data integration and analytics have significantly increased.
- Furthermore, core data elements from disparate information systems have been harmonized and are well understood.

Data Is Worthless If You Don’t Communicate It

- To often we compile vast troves of data and analytical reports that never see the light of day
- We need to communicate findings well, rather than assuming the results speak for themselves
- Simple communication model
  - Understanding of problem
  - Measurement of impact of problem
  - What data is available
  - Initial solution hypothesis
  - The Solution
  - The impact of the Solution



Decisions Don’t Start with Data



- Data can provide new insights and evidence to inform our toughest decisions, but numbers alone won’t convince others.
- Good stories – with a few key facts woven in – are what attach emotions to your argument, prompt people into unconscious decision making, and ultimately moves them to action.

Continuing Education

Let’s Talk Informatics has been certified for continuing education credits by;

- College of Family Physicians of Canada and the Nova Scotia Chapter for 1 Mainpro+ credit.
- Digital Health Canada for 1CE hour for each presentation attended. Attendees can track their continuing education hours through the HIMSS online tracking certification application, which is linked to their HIMSS account.

THANK YOU

Question ?