

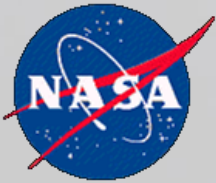


Data Analytics to Solve the Unique Challenges of Astronaut Medical Data

Jacqueline Charvat, Ph.D.

Mary Wear, Ph.D.

Mary Van Baalen, Ph.D.

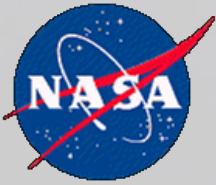


- **Aim of LSAH is:**

- Describe morbidity (acute and chronic) and mortality of astronauts
- Determine whether unique occupational exposures encountered by astronauts are associated with increased risks of morbidity or mortality
- Repository for all archived medical data
 - Provide data for internal and external requestors

- **Exposures**

- **Spaceflight**
 - Microgravity
 - Chemicals
 - Radiation
 - Closed loop life support system
 - Particulates
 - Lasers
 - Sunlight
- **High Performance Jet Training**
- **Analog Training**
 - NEEMO
 - Antarctica



Lifetime of Exposures

Spaceflight Exposures

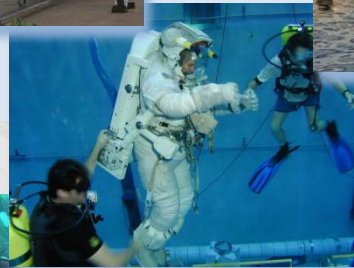
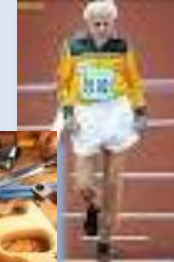


Pre-NASA Exposures

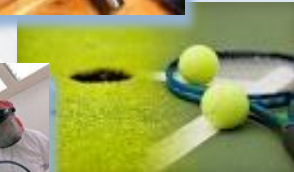
ASCAN Training

Active Astronaut

Retirement



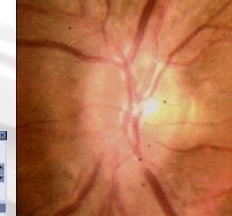
Terrestrial Exposures



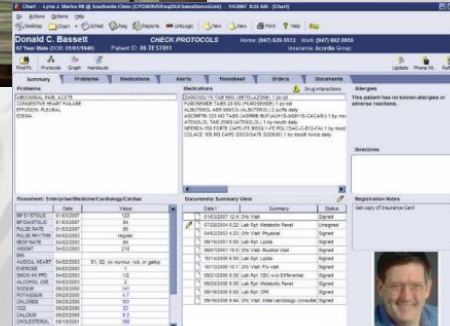


Challenges of Astronaut Data

- Celebrity Effect
- Selection Bias and Healthy worker effect
- Cohort Effect
- Screening Bias
- Small N
 - Varied exposure to spaceflight
- Over-interpretation



OD right	AL: 23.52 mm (SNR = 223.8) K1: 43.38 D / 7.78 mm @ 9° K2: 43.95 D / 7.68 mm @ 90° R / SE: 7.73 mm (SD = 43.67 mm) Cyl: -0.27 D @ 9° opt. ACD: 3.52 mm	OS left	AL: 23.80 mm (SNR = 113.8) K1: 42.77 D / 7.71 mm @ 1° K2: 44.12 D / 7.65 mm @ 91° R / SE: 7.68 mm (SD = 43.95 mm) Cyl: -0.39 D @ 1° opt. ACD: 3.67 mm
	Eye Status: phakic		
Alcon SN60WF	MTA4U0	Alcon SN60WF	MTA4U0
SP: 1.84	SP: 0.17	SP: 1.84	SP: 0.17
100 (D) REF (D)	100 (D) REF (D)	100 (D) REF (D)	100 (D) REF (D)
22.5 -0.26	19.0 -1.16	21.0 -0.78	18.0 -1.12
21.5 -0.26	18.5 -0.28	20.0 -0.44	17.5 -0.55
21.0 0.07	18.0 0.10	20.0 -0.11	17.0 0.03
20.5 0.41	17.0 0.48	19.5 0.22	16.5 0.40
20.0 0.73	17.0 0.85	19.0 0.55	16.0 0.77
19.5 1.04	16.5 1.22	18.5 0.87	15.5 1.13
Errors: 100: 23.11	Errors: 100: 18.13	Errors: 100: 19.93	Errors: 100: 17.03
SN6ADI RESTOR	SN6ATJ4 & 5	SN6ADI RESTOR	SN6ATJ4 & 5



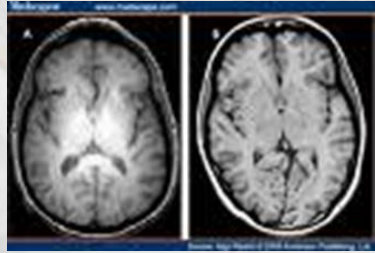
Data Compilation & Audit

- Patient Medical Records
 - EMR
 - Hardcopy records
- Mission Medical Repository
 - Monitoring around spaceflight missions
- Individual Exposure Profile
- Training Records (NBL, survival, NOLS, etc.)
- Spacecraft Environmental Data
- Biospecimen Data
- Research Study Data
- "Other" Data

3 Tesla MRI NASA Intracranial Pressure Protocol
 Technique:
 3 Tesla MRI of the brain was obtained using isovolumetric T1 and T2 images of the whole brain. High resolution fat saturated, isovolumetric axial and oblique sagittal images of the orbits were obtained, along with axial FLAIR images of the brain.
 Indication: Flight Evaluation

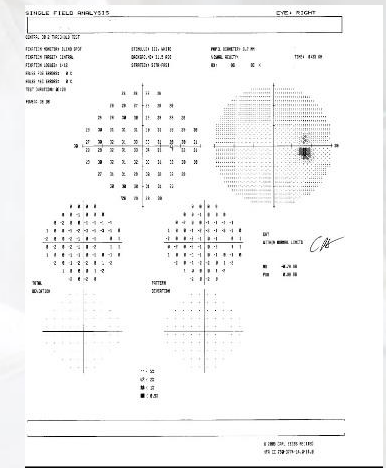
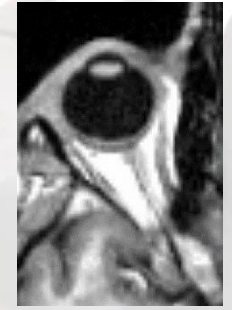
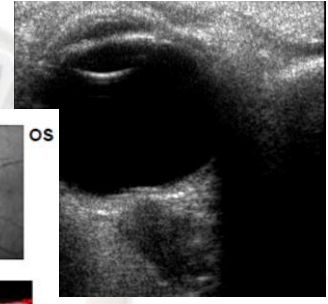
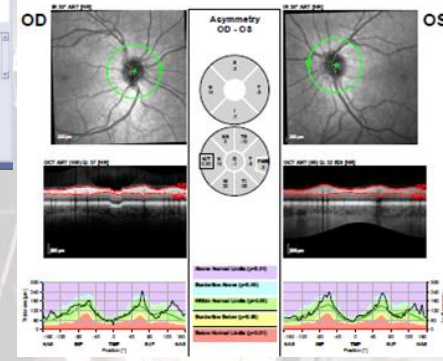
Findings:
 Brain:
 The ventricles and sulci are normal for the patient's age. The cisterns are within normal limits.
 The sella turcica shows a partially empty sella configuration, without chiasmatic herniation.
 No evidence of intracranial lesions. A non-specific punctuate FLAIR hyperintensity is noted in the deep white matter of the right frontal lobe, unchanged since 2011.
 Perivascular space visualization is as expected for the patient's age. No extra-axial fluid collections are seen.
 The paranasal sinuses are clear.

ORBITAL ANALYSIS:
 Right eye
 Left eye
 Optic Nerve sheath diameter (ONSD) in cm axial 0.68 cm
 0.75 cm
 Optic Nerve sheath diameter (ONSD) in cm sagittal 0.68 cm
 0.74 cm

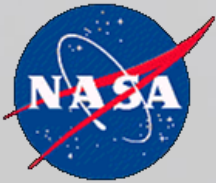


Optic Nerve Diameter (OND) in cm axial 0.31 cm
 0.31 cm
 Optic Nerve Diameter (OND) in cm sagittal 0.30 cm
 0.39cm
 Optic Nerve sheath to nerve ratio (ON SNR) 2.19 ? 2.27
 2.42 ? 1.90
 Papilledema not present
 Globe flattening grade 0-3 0
 Optic nerve tortuosity grade 0-3 0
 Optic Nerve hyperintensity 0-3 1

Impression:
 1. Normal MRI of the brain without contrast.
 2. No nerve sheath distention, optic globe flattening, or optic nerve hyperintensity to suggest increased intracranial pressure.
 3. Normal sella.

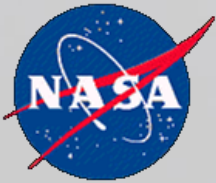


11-28-05 pt picked copy of tri. Rx (20)
 5-3-08 pt. 2 pt c/o smudge in vision x weeks. not always there but in the center when it is. had exam ophthalm. in AZ. told of some wrinkling sched tomorrow Doc/oc (R)



Complicating Factors: Data Harmonization

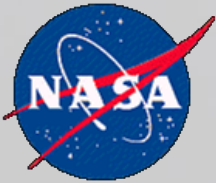
- Multiple sources collected for multiple reasons collected in multiple formats
 - Data entry into different places
- Dates
 - **Flight Day** ≠ **GMT** ≠ **MET** ≠ **Date of Entry**



IMPALA OVERVIEW

**“Information Management Platform
for Analytics & Aggregation”**

- Platform to store, process, perform and share visualizations and advanced analytics of astronaut data to solve our data challenges
- Analyze multi-structured data, gain insights and make data driven decisions
- Self-service system to visualize large datasets.
- Share deep insights by collaborating in (near) real time across the board



IMPALA: Trifacta

Trifacta - Transformer

Customers ▾ Generate Results

Full Datasource - 444.49KB ▾ 27 Columns 2,208 Rows 5 Data Types Filter in grid

id	segment	ownHomeFlag	contactPreference	leadSource	dateOfBirth	#	salesClass
1	Services	N	phone	partner	12/4/89	3	M
2	is	N	donotcall	Twitter	11/18/94	2	S
3		N	mail	Twitter	7/16/94	4	M
4	is	N	email	Facebook	6/8/85	1	S
5	is	N	facebook	Twitter	7/31/95	2	S
6	is	Y	donotcall	List - Purchase	3/29/91	2	S
7		N	phone	Outbound-	4/17/93	5	S
8		N	mail	LinkedIn	10/26/85	2	Married
9		N	mail	LinkedIn	1/23/89	4	M
10		N	donotcall	Referral	7/2/88	2	S
11		N	phone	Web - Activity	6/22/90	2	M
12	Services	N	facebook	Other	8/31/88	1	S
13		Y	none	Twitter	2/7/90	3	Single
14		N	none	LinkedIn	3/29/89	3	Single
15	Services	N	donotcall	Google	3/16/83	5	M
16	Services	N	none	Outbound-	7/2/88	5	M

TRANSFORM EDITOR



Enter transform expression

Add to Script

Mock data



Complicating Factors: Metadata

- Who, what, why about our data
- Different types of data = different types of metadata
- Documentation is static, outdated or doesn't exist
- Historical context and description of changes across time largely absent
- Granularity differs across data types and across time
- Tags to find appropriate data largely absent



IMPALA: Waterline

WATERLINE DATA INVENTORY Dashboard Browse Manage restaurant Advanced Search

Search Results
Criteria: Content Type: CSV Edit
Filters:

1. Business metadata-based search

2. Discover meaning of fields and tag automatically

3. Automatically propagate user-defined tags (crowdsource ontology)

4. Automated facet creation based on metadata

5. Multi-faceted drill down

Tags

- Address.Borough (17)
- Food Service.Alcohol Licenses.Code (2)
- Food Service.Cuisine (21)
- Food Service.Inspections.Grade (1)
- Food Service.Inspections.Violation Description (10)
- Social Security Number (4)
- Social Security Number No Separators (9)

Sort by:

Name

- Topic.Food Service (2)
- Phone Number (0)
- Occupation (0)
- US Address (0)
- US State Abbreviation (0)
- Global City (0)
- People Names (0)
- Salutation (0)
- First Name (0)
- US County (0)

Less Clear

Field Name	Value	Count	Percentage
Social Security Number	100%	1	100%
Food Service.Inspections.Violation Description	100%	1	100%
Food Service.Inspections.Grade	100%	1	100%
US City	43%	2	100%
Address.Borough	92%	1	100%
Food Service.Cuisine	84%	1	100%
Address.Borough	92%	1	100%
Address.Borough	92%	1	100%
Address.Borough	92%	1	100%
Social Security Number	88%	1	11%
Social Security Number	86%	1	100%
Social Security Number	86%	1	100%
Food Service.Alcohol Licenses.Code	84%	1	100%
Address.Borough	84%	1	100%
Address.Borough	84%	1	100%
Address.Borough	84%	1	100%
Food Service.Cuisine	77%	1	100%
Food Service.Cuisine	77%	1	100%

Mock data

1 - 66 of 66 items

CUISINE DESCRIPTION
Info Profile Values

Most Frequent Values

- Pizza/Italian 1054
- Hamburgers 855
- Chicken 740
- Ice Cream, Gelato, Yogurt, Ices 724
- French 437
- Asian 440
- Delicatessen 419
- Jewish/Kosher 342
- Indian 338
- ... 322
- ... 318
- ... 315
- ... 313



Complicating Factors: Data Analytics and Visualizations

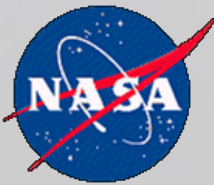


THE
POWER
TO KNOW.

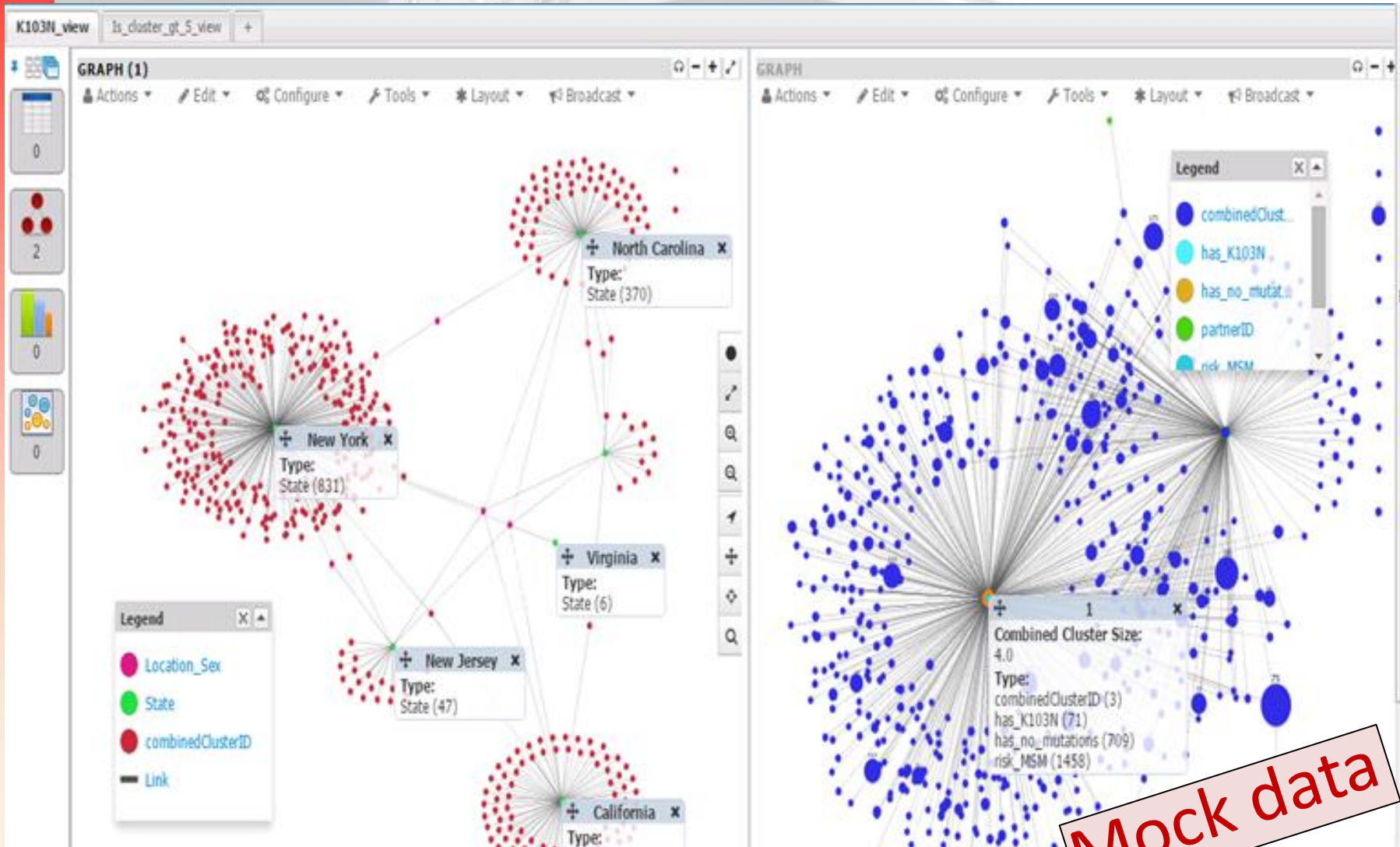


```
title 'QTc Change from Baseline by Week and Treatment';
proc sgplot data=QTcBand;
  format week qtcweek.;
  styleattrs datalinepatterns=(solid);
  band x=wk lower=L0 upper=L30 / fill legendlabel='Normal'
      fillattrs=(color=white transparency=0.6);
  band x=wk lower=L30 upper=L60 / fill legendlabel='Concern'
      fillattrs=(color=lightgray transparency=0.6);
  band x=wk lower=L60 upper=L90 / fill legendlabel='High'
      fillattrs=(color=gray transparency=0.6);
  vbox qtc / category=week group=drug groupdisplay=cluster nofill;
  scatter x=wk y=QTc / group=drug name='a' nomissinggroup;
  text x=wk y=ylabel text=label / contributeoffsets=none;
  xaxistable risk / class=drug colorgroup=drug location=inside;
  refline 26 / axis=x;
  xaxis type=linear values=(1 2 4 8 12 16 20 24 28) valueshint
      min=1 max=29 display=(nolabel)
      colorbands=odd colorbandsattrs=(transparency=1);
  yaxis label='QTc change from baseline' values=(-120 to 90 by 30);
  keylegend 'a' / title='Treatment:' linelength=20;
run;
```

```
proc genmod data=resp descend;
  class id treatment(ref="P") center(ref="1") sex(ref="M")
      baseline(ref="0") / param=ref;
  model outcome=treatment center sex age baseline / dist=bin;
  repeated subject=id(center) / corr=unstr corrw;
run;
```



IMPALA: Visualizations



Mock data



IMPALA Workflow Alpine

Native integration with R & Python scripts & executables

The screenshot shows the Alpine interface for a workflow named "Account Churn". The workflow is a directed graph of tasks: "account_statements" and "accounts" feed into "Table join - Account Info and Statements"; "zip_income" feeds into "Table join - Accounts and Zips"; "Note-Data" feeds into "Note-Transformations"; "Table join - Account Info and Statements" and "Table join - Accounts and Zips" feed into "Aggregate By Account"; "Note-Sampling" feeds into "Random Sampling"; "Note-Transformations" feeds into "Random Sampling"; "Aggregate By Account" and "Random Sampling" feed into "Train - 80%"; "Train - 80%" feeds into "Decision Tree", "Logistic Regression", and "Naive Bayes"; "Train - 80%" and "Test - 20%" feed into "ROC"; "Decision Tree", "Logistic Regression", and "Naive Bayes" feed into "Note-Models". The right sidebar shows actions like "Run Now", "Stop Running", "Add a note", "Add/Edit tags", "Rename", "Copy latest version", and "Delete work file". An activity log shows "Workflow Account Churn finished running." and "Alpine Data Labs updated workflow Account Churn Changed where docs are stored to".

Advanced Analytics through drag & drop interface

Collaborate with team through a dedicated workspace

The screenshot shows the Alpine interface for a workspace named "Quantum Demo Workspace". The workspace details include: Owner: [redacted], Status: On Track, Target: N/A (requires milestones), Progress: 0 / 0, and Show: All Activity | Insights. A list of recent activities is shown: "[redacted] added [redacted] and 2 others" (20 hours ago), "[redacted] updated workflow Exploration Configured correct data set" (20 hours ago), and "[redacted] added file Exploration" (20 hours ago). The right sidebar shows actions like "Add a note", "Add an insight", "Edit Workspace", "Add or Edit Members", "Add a sandbox", and "Delete this Workspace".



Complicating Factors: Text Data

01/31/2013 9:41 AM	PMC: Pre-PMC 5.0 zero-out
01/23/2013 2:21 PM	PMC: Pre-PMC 4.0 Test#11

Doc ID: 158

Properties: Private Medical Conference at FMClinic on 02/06/2013 11:45 AM by Richard

Start GMT: 11:00
End GMT: 11:20
PMC Duration: 0:20

History of Present Illness:

History from: Crewmember

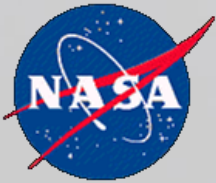
Chief Complaint: congestion

HPI: Back pain improved but not gone. Using Ibuprofen as needed (1-2/day). Visual changes have worsened slightly. Now also noticing sinus and head congestion - probably secondary to microgravity fluid shifting. Will add Nasonex spray for sinus congestion. Workload for ATV docking preparations this week have resulted in not sleeping enough - used Provigil this morning to help alertness.

For Help, press F1

242 Vitamin	Routine	Nutritional Supplementa	Multivitamin	9999	9999	Oral
243 Fiber Supplement	Periodic	Fiber Supplement	Psyllium (Metamucil)	9999	9999	Oral
244 Antilipemic	Routine	Hyperlipidemia	Atorvastatin (Lipitor)	9999	9999	Oral
245 Decongestant	Periodic	Other	Oxymetazoline (Afrin Nasal Spr.	9999	9999	Nasal
246 Antihypertensive	Routine	Hypertension	Lisinopril (Zestril)	9999	9999	Oral
248 Analgesic	Periodic	Pain	Ibuprofen (Motrin)	9999	9999	Oral
249 Antibiotic/Corticosteroid	Periodic	Irritation	Tobramycin/Dexamethasone (T	9999	9999	Optic
250 Lubricant	Periodic	Irritation	Carboxymethylcellulose OPHTH .5	%		Optic
251 Analgesic	Periodic	Pain	Acetaminophen (Tylenol)	325	mg	Oral
252 Antibiotic/Corticosteroid	Periodic	Irritation	Tobramycin/Dexamethasone (T	9999	9999	Optic
253 Vitamin	Routine	Nutritional Supplementa	Multivitamin	9999	9999	Oral
254 Analgesic	Periodic	Congestion	Ibuprofen (Motrin)	600	mg	Oral
255 Decongestant	Periodic	Congestion	Oxymetazoline (Afrin Nasal Spr.	9999	9999	N
256 Antibiotic	Periodic	Irritation	Ciprofloxacin OPHTH Oint (Cilo	9999		
257 Antibiotic	Periodic	Irritation	Ciprofloxacin OPHTH Soln (C			

Mock data



Complicating Factors: Text Data

"CM WORE HIS PRIME TRAINING GLOVES AND REPORTED '**SORE**' **FINGERTIPS**. THE SIZING OF THE GLOVES WAS REPORTED AS ACCEPTABLE, WITH NO RESIZING REQUESTS. FOR THE TIME BEING, CM REQUESTED TO **REMOVE THE NAIL HARDENER** FROM HIS CREW PREFERENCE ITEMS. ALL OTHER SUIT FIT COMMENTS WERE FAVORABLE. "

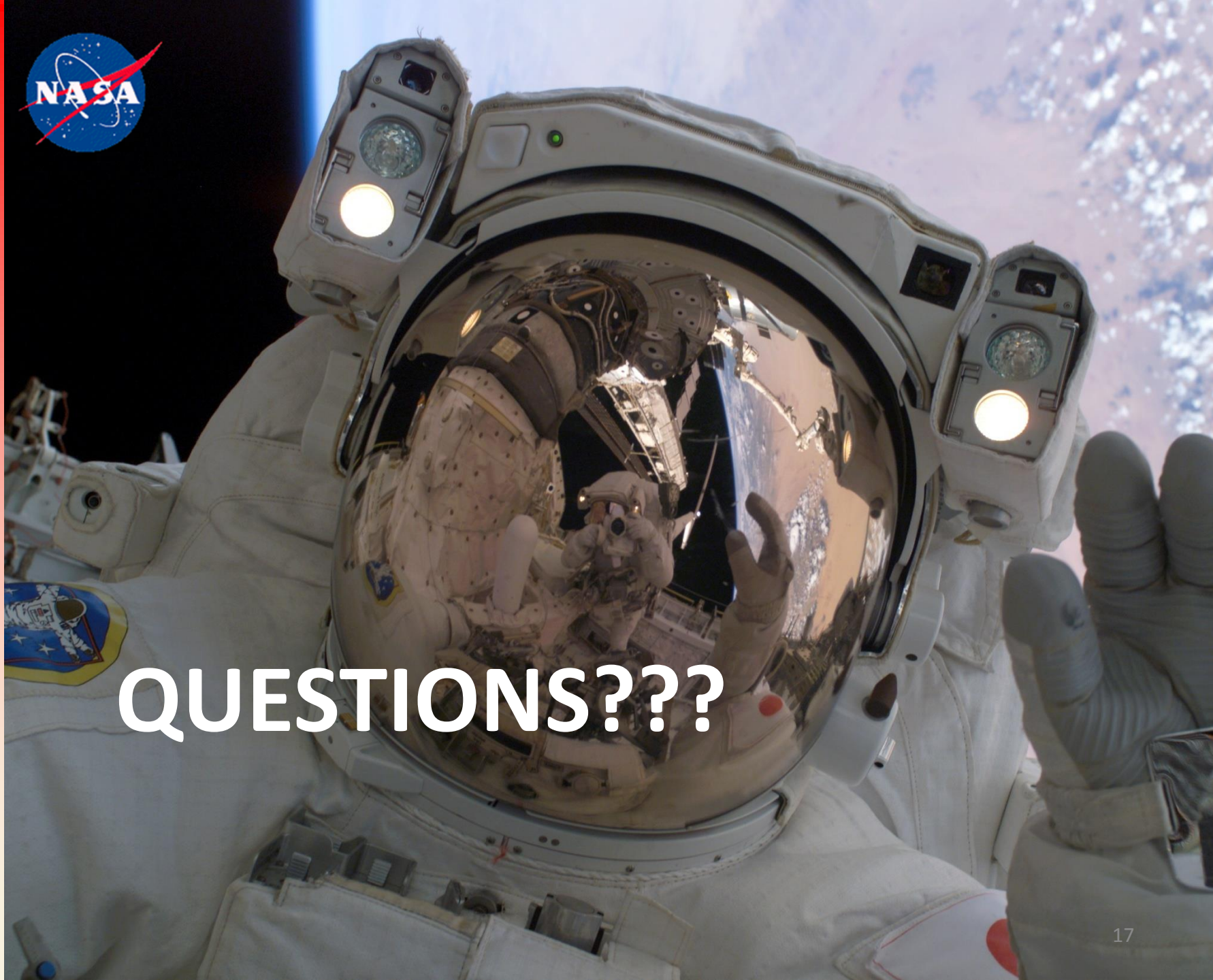
**Over 3000 pages of Sizing Comments
Unable to be Analyzed**

Mock data



Summary

- Astronaut data and analytics of astronaut data is complex
- Creative and unique solutions are needed and some are currently being implemented
- Solutions for qualitative analysis including NLP are still needed



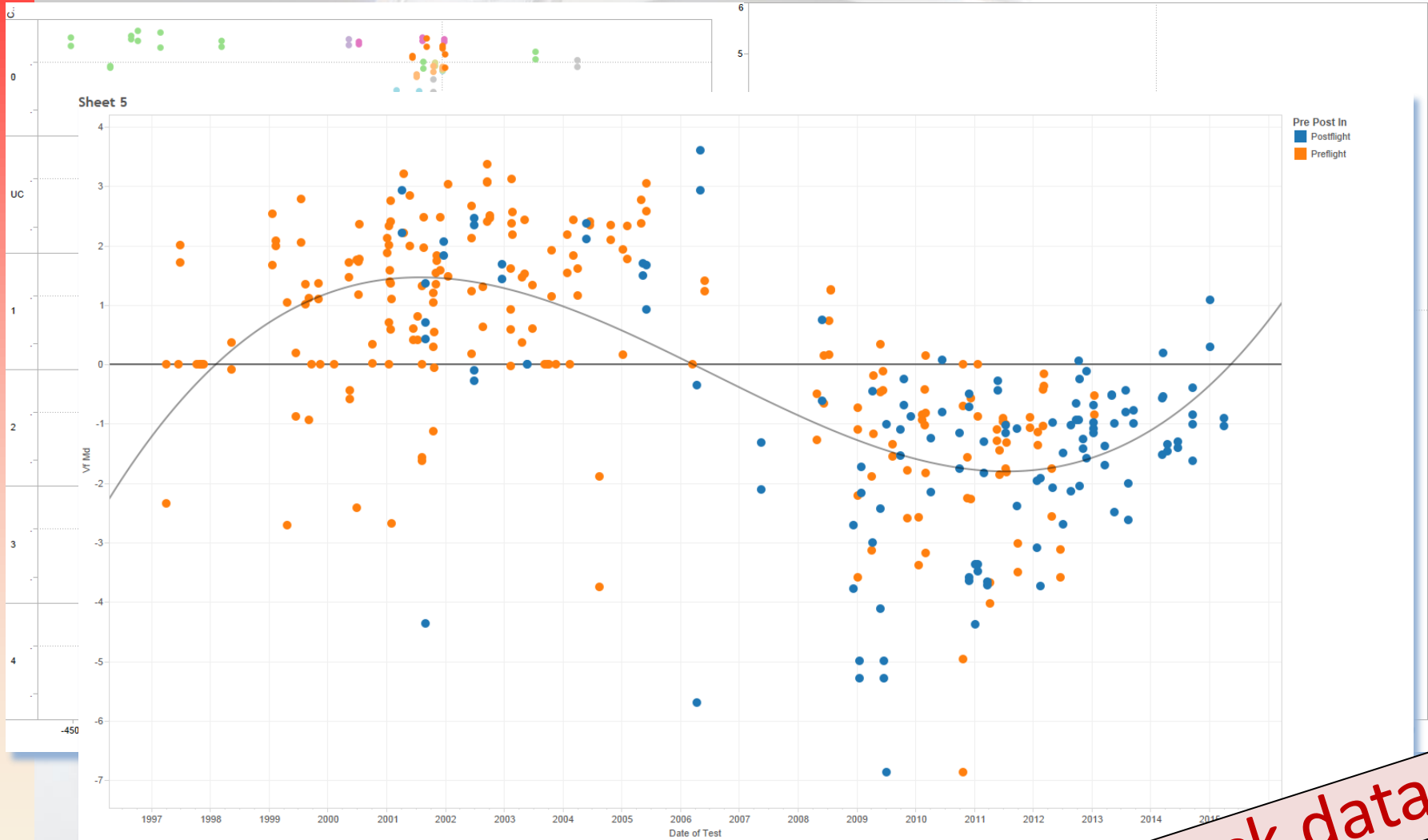
QUESTIONS???

Back Up





Trending and Analyses



Date of Test vs. VfM/d. Color shows details about Pre Post In. Details are shown for ID. The data is filtered on Missing Here, which keeps 0 and 1.

Mock data