splunk> .conf2017

Go Splunk Yourself

How Using Splunk to Analyze My Biometric Data Has Improved My Quality of Life

Josef Kuepker | Staff Sales Engineer

27 September 2017 | Washington, D.C.

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

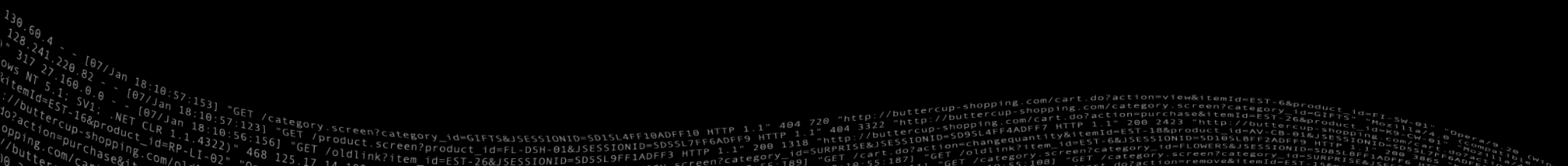
The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

splunk> .conf2017

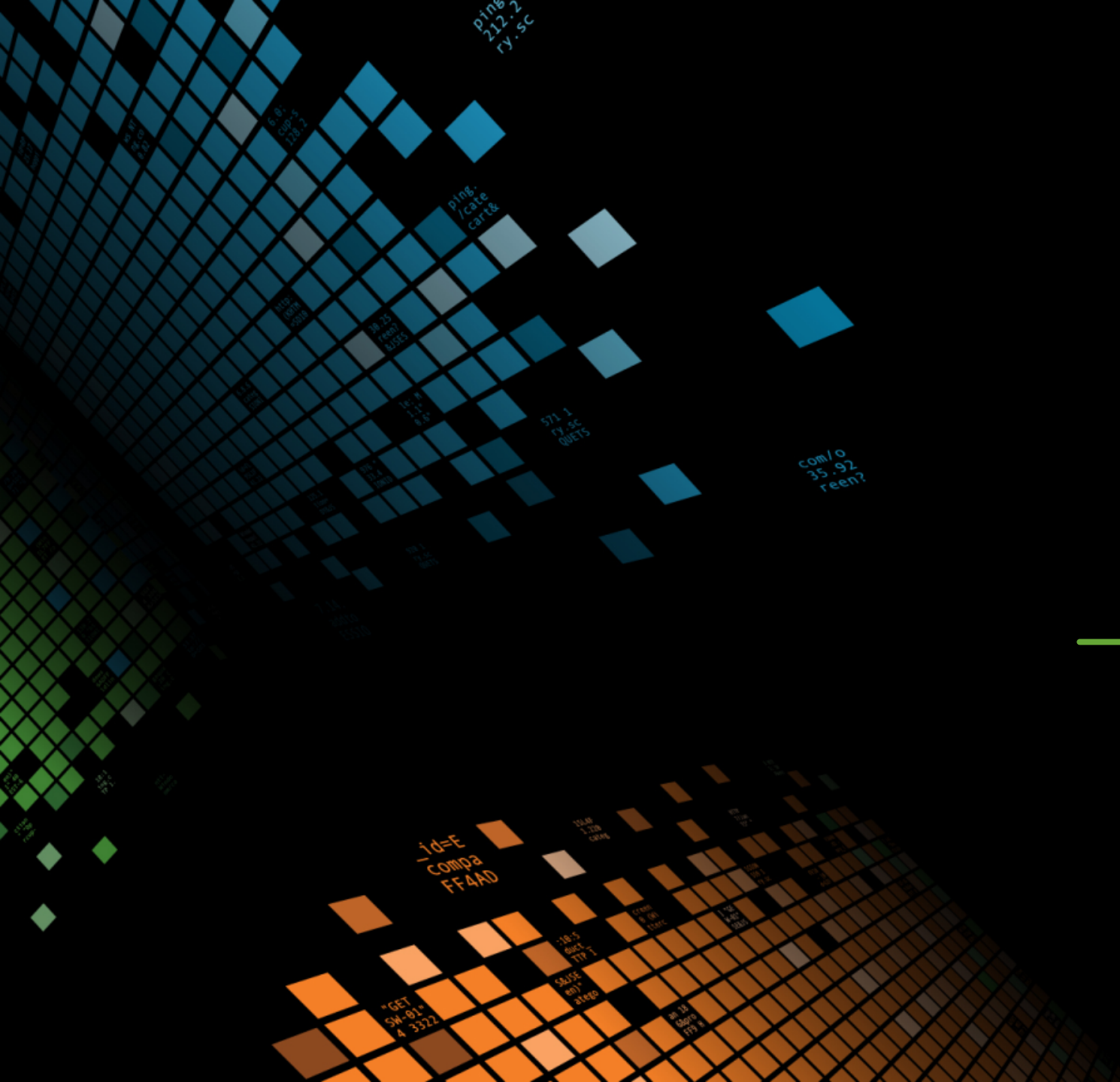
Agenda

What are we learning today?

- About me
- What is Sleep Apnea?
- Why I started this project
- My diagnosis and timeline
- Data and analysis
- Future state



splunk> .conf2017



About Me

Who is this guy?

Who Is This Guy?

- Josef Kuepker
- Splunk since mid-2016
- Previous Lives
  - JP Morgan Chase
    - Incident Response, Cybersecurity Engineering
  - Department of Defense
    - Incident Response, Information Assurance
- Hobbies
  - Travel (Iceland is my favorite.)
  - Tinkering (Splunk, Gadgets, etc.)

Why Did I Start This Project?

And what is sleep apnea?

What Started This Project?

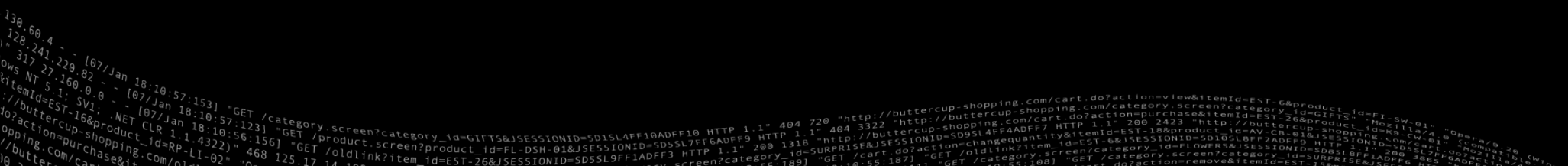
How I got here

E-Mail

To: [redacted]

Subject: A secure message from your provider's office

Message: Hi Josef, yoursleep study does show sleep apnea. You stop your breathing 46 times per hour during sleep. I have submitted a request to your insurance to do a cpap test in the lab where you will sleep in the office and we will put the air mask on your nose to make the apnea go away. Once it is approved [redacted] will call you to schedule it.

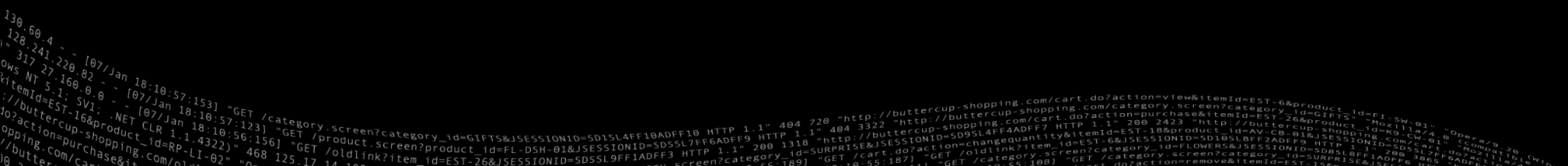


splunk> .conf2017

Sleep Apnea

The reason I am here

- A common, chronic sleep disorder identified by pauses in breathing or shallow breathing, which can:
  - Last from seconds to minutes
  - Vary the amount of light to deep sleep
  - Lead to poor sleep quality and excessive daytime sleepiness



My Diagnosis Timeline

How long did this take?

December 2016

Sleep Study in lab and diagnosis of 46 pauses in breathing per minute

January 2017

Sleep Study in lab with CPAP to determine the best pressure (8 cmH2O)

February 2017

Research of CPAP and arguing with DME provider

April 2017

Receive the correct CPAP device and started Splunking it

splunk> .conf2017

Getting The Data and Analyzing It

Challenges and Findings

splunk> .conf2017

Getting The Data

What data is available?

- Depends on your device
  - Research. Not all devices are equal. (http://www.apneaboard.com/wiki/index.php?title=Archangle:Machine_Choices)
  - Data is generally written in a binary format called EDF/EDF+.
- Will need conversion for Splunk.
  - Several Python libraries available.
  - SleepyHead is a free tool that can import the EDF(+) and export to CSV.

RESMED AutoSet AirSense 10

EDF+ Log Format

What does the original data look like?

The header record contains

8 ascii : version of this data format (0)	0
80 ascii : local patient identification	MCH-0234567 F 02-MAY-1951 Haagse_Harry
80 ascii : local recording identification.	Startdate 02-MAR-2002 EMG561 BK/JOP Sony. MNC R Median Nerve.
8 ascii : startdate of recording (dd.mm.yy)	17.04.01
8 ascii : starttime of recording (hh.mm.ss).	11.25.00
8 ascii : number of bytes in header record	768
44 ascii : reserved	EDF+D
8 ascii : number of data records (-1 if unknown)	2
8 ascii : duration of a data record, in seconds	0.050
4 ascii : number of signals (ns) in data record	2

	1st signal	2nd signal
ns * 16 ascii : ns * label	R APB	EDF Annotations
ns * 80 ascii : ns * transducer type (e.g. AgAgCl electrode)	AgAgCl electrodes	
ns * 8 ascii : ns * physical dimension (e.g. uV)	mV	
ns * 8 ascii : ns * physical minimum (e.g. -500 or 34)	-100	-1
ns * 8 ascii : ns * physical maximum (e.g. 500 or 40)	100	1
ns * 8 ascii : ns * digital minimum (e.g. -2048)	-2048	-32768
ns * 8 ascii : ns * digital maximum (e.g. 2047)	2047	32767
ns * 80 ascii : ns * prefiltering (e.g. HP:0.1Hz LP:75Hz)	HP:3Hz LP:20kHz	
ns * 8 ascii : ns * nr of samples in each data record	1000	60
ns * 32 ascii : ns * reserved		

Log Format Conversion

Binary → ASCII

```
    n = f.signals_in_file - 1
    for i in np.arange(n):  # for each signal but the last)
        y = 0
        sample_dur = f.file_duration / f.datarecords_in_file
        time_inc = sample_dur / f.getSampleFrequency(i)
        epoch = int(time.mktime(
            time.strptime(str(f.getStartdatetime()), '%Y-%m-%d %H:%M:%S')))  # generate starting epoch for each sample
        for x in f.readSignal(i):  # for each sample in the signal
            y += time_inc
            result = ""
            result += datetime.datetime.fromtimestamp(float(epoch) + y / 60.0).strftime('%m-%d-%Y %H:%M:%S.%f')
            result += " [" + str(i) + ":" + f.getLabel(i).replace(".", "_") + "] " + str(x) + " " + str(
                f.getPhysicalDimension(i))
            result += " " + str(f.getPhysicalMinimum(i)) + ":" + str(f.getPhysicalMaximum(i))
            result += " " + str(f.getDigitalMinimum(i)) + ":" + str(f.getDigitalMaximum(i))
            result += " " + str(f.getFileDuration())
            result += " " + str(f.getSampleFrequency(i)) + "\n"
# Older log formats.
            #result += "  [signal=" + str(i) + "]"
            #result += " patientName=\"" + f.getPatientName() + "\""
            #result += " patientGender=\"" + f.getGender() + "\""
            #result += " patientBirthday=\"" + f.getBirthdate() + "\""
            #result += " recording=\"" + f.getPatientAdditional() + "\""
            #result += " patientCode=\"" + f.getPatientCode() + "\""
            #result += " adminCode=\"" + f.getAdmincode() + "\""
            #result += " technician=\"" + f.getTechnician() + "\""
            #result += " equipment=\"" + f.getEquipment() + "\""
            #result += " recordinAdditional=\"" + f.getRecordingAdditional() + "\""
            #result += " phyMax=\"" + str(f.getPhysicalMaximum(i)) + "\""
            #result += " phyMinMax=\"" + str(f.getPhysicalMinimum(i)) + ":" + str(f.getPhysicalMaximum(i)) + "\""
            #result += " fileDurSec=\"" + str(f.getFileDuration()) + "\""
            #result += " phyMin=\"" + str(f.getPhysicalMinimum(i)) + "\""
            #result += " digMax=\"" + str(f.getDigitalMaximum(i)) + "\""
            #result += " digMin=\"" + str(f.getDigitalMinimum(i)) + "\""
            #result += " digMinMax=\"" + str(f.getDigitalMinimum(i)) + ":" + str(f.getDigitalMaximum(i)) + "\""
            #result += " phyDim=\"" + str(f.getPhysicalDimension(i)) + "\""
            #result += " prefilter=\"" + str(f.getPrefilter(i)) + "\""
            #result += " transducer=\"" + str(f.getTransducer(i)) + "\""
            #result += " samFreq=\"" + str(f.getSampleFrequency(i)) + "\""
            #result += " " + f.getLabel(i).replace(" ", "_") + "=\"" + str(x) + " " + str(f.getPhysicalDimension(i)) + "\"\n"
            sendOut.write(result)
```

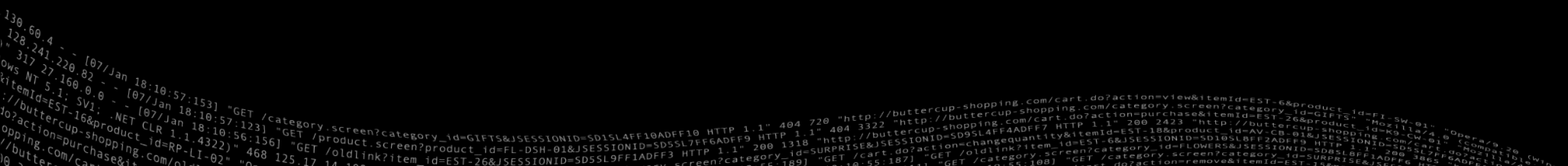
PEP 8: block comment should start with '# '

splunk> .conf2017

Converted Log Format

How changing the format helped

- REV 1 (~39GB):
  - 07-18-2017 06:05:58.960000 [signal=0] phyMax="3.0" fileDurationSecs="16380" phyMin="-2.0" digMax="1500" digMin="-1000" phyDim="L/s" samFreq="25.0" Flow_40ms="0.232”
- REV 2(~30GB):
  - 07-18-2017 06:05:58.960000 [signal=0] phyMinMax="-2.0:3.0" fileDurSec="16380" digMinMax="-1000:1500" samFreq="25.0" Flow_40ms="0.232 L/s”
- REV 3 (~9GB):
  - 07-18-2017 06:05:58.960000 [0:Flow_40ms] 0.232 L/s -2.0:3.0 -1000:1500 16380 25.0



splunk> .conf2017

What Does The Data Tell Me?

How does travel affect my sleep?

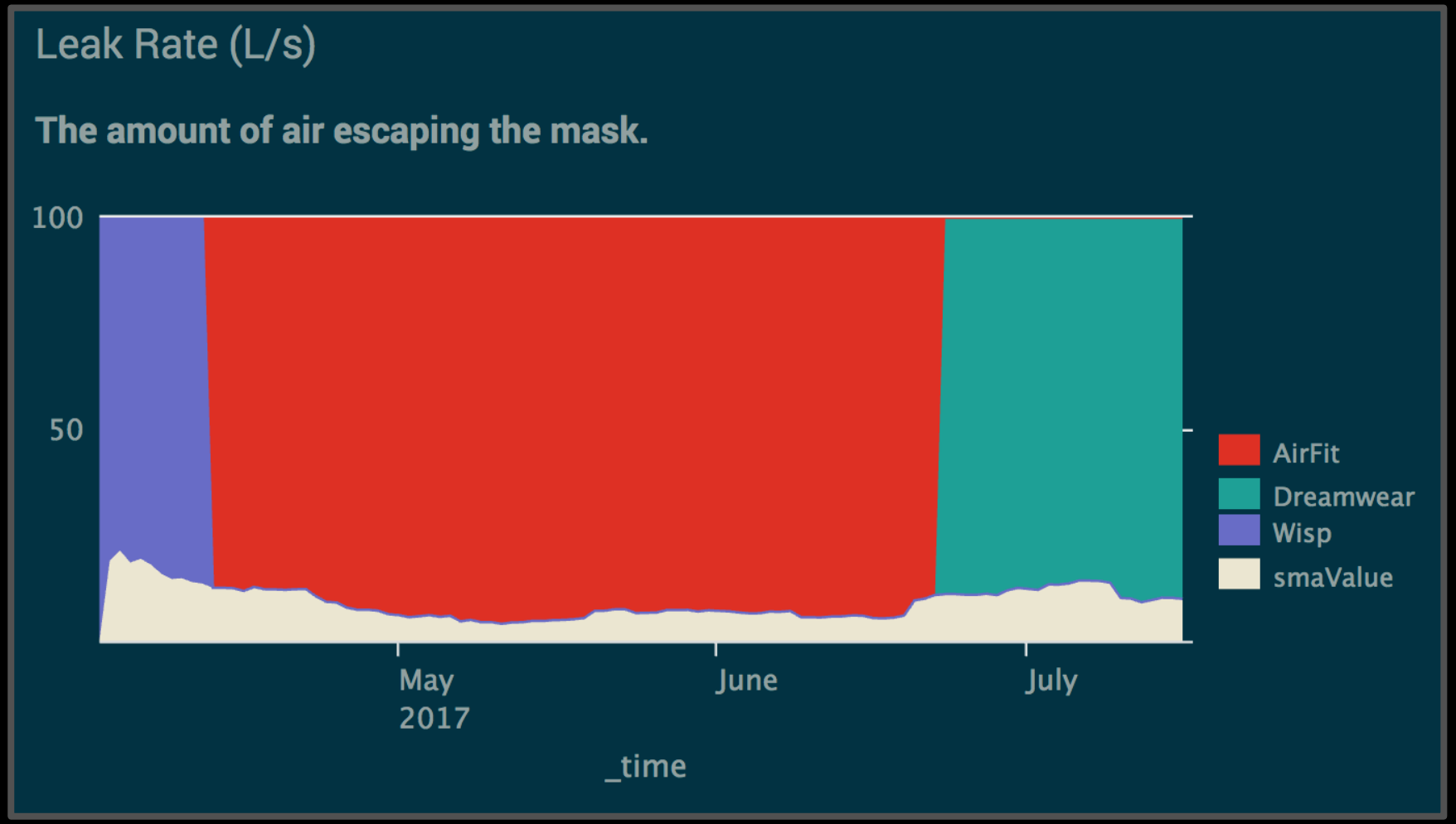
- My sleep quality is not really affected by travel.
- I am less likely to use my machine.
- I travel a lot!

splunk> .conf2017

What Does The Data Tell Me?

Does mask choice make a difference?

- The AirFit seems to be the best overall mask for me.
- Leak Rate was the only metric that had any significant change.



More Metrics!

- 97/105 nights
- 6.27 avgHrs/night
- 660mL per inhale
  - Avg human is 500mL
- 13.77 bpm
  - 12<X<25 is good
- Extremely low leak rate

splunk> .conf2017

What Did I Learn?

Takeaways

1. Mask choice matters to a point, but pick for comfort.
2. Travel doesn't have a significant impact on my sleep patterns.
3. There is a long timeline from suspicion to treatment. Be proactive and call frequently.
4. Be persistent with your DME and have your provider help.
5. I don't get enough sleep.

What's Next?

1. Work on converting the annotations EDF+ file. Slightly different format.
2. Automate the process with RPi and wireless SD card.
3. Find additional data sources to add (sleep journal, temperature, etc).
4. Switch my device to Auto Mode to see if a variable pressure is better/worse.

THANK YOU!

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017