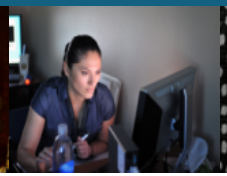
SAND2019-5212C

Sandia National Laboratories

Visualizing a Safety Model with R Shiny



PRESENTED BY

Adrian Perez & Scarlett Marklin

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

325 Safety Incidents

Bureau of Labor Statistic rate of 2.8 cases per 100 full-time workers for 13,000 SNL employees.

Who Are We?

Adrian Perez - jespere@sandia.gov

Education

- MS in Applied Statistics
- Programming with R for 10 years now

Sandia National Labs

- Data Scientist (July 2018)
- Data Sciences Organization (10734)
  - Productization of Machine Learning

Wrote 4714 lines of R code for this project!

Scarlett Marklin – smarkli@sandia.gov

Education

- MS in Physics, Demography, Geography, and Geology

Sandia National Labs

- Computer Scientist (Jan 2018)
- Assurance Data Sciences (9221)

First Major Project using R!

Assurance for Safety

A tool for uncovering risk

Something That Can…

1. Inform – direct attention across the labs, answer question(s)
2. Anticipate – find predictors that have a strong relationship to safety, promoting a proactive stance
3. Be Proactive - Be put into a visualization that users can interact with to drive decision making
4. Prevent – provide leading indicators that could focus attention and monitor risk

Why a Behavioral Model?

Safety has not only a systems component, but a largely behavioral component as well

It's great for answering

- What set of predictors in combination are effective in predicting the likelihood of safety incidents
- Interpretability of indicators
- Highlights areas for management to focus

Expected benefits

- Identifying leading indicators

Building the Model

Why a Behavioral Model? - REPEAT TITLE

Allowed for Unique behavioral predictive model to address two goals:

1. Identify organizations at high-risk for one or more safety incidents in the next 6 months after the end of each quarter.
2. Identify predictive indicators from corporate data systems that demonstrate significant relationships to future safety incidents.

Unique because it was developed using Sandia data and Sandia safety incidents – customized for Sandia!

The Safety Model

Does give risk predictions that will point to the most likely outcome for an individual

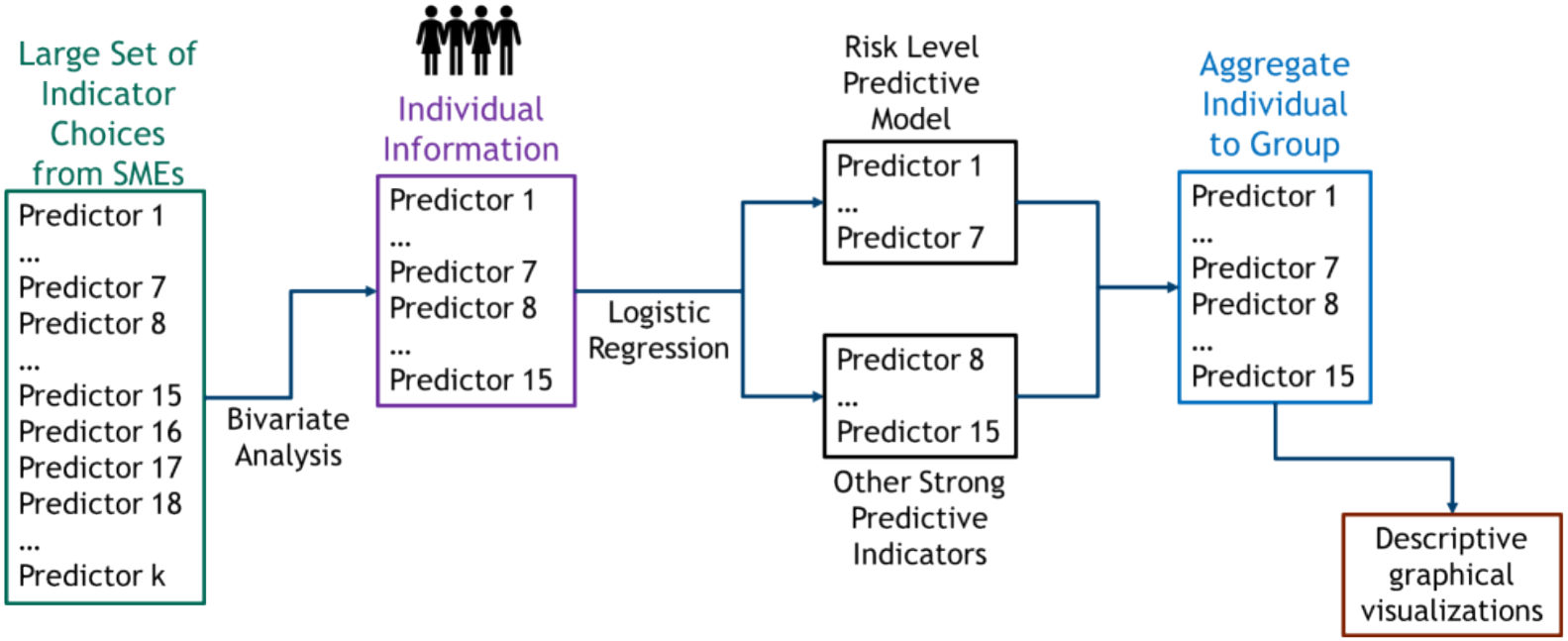
- As an employee who will have an incident in the next six months.
- As an employee who will not have an incident in the next six months.

Does not identify a specific person who will have an incident at a specific time.

Predictive indicators are not causal.

- They exhibit a demonstrated relationship to future incidents.
- They give early warning that can be investigated.

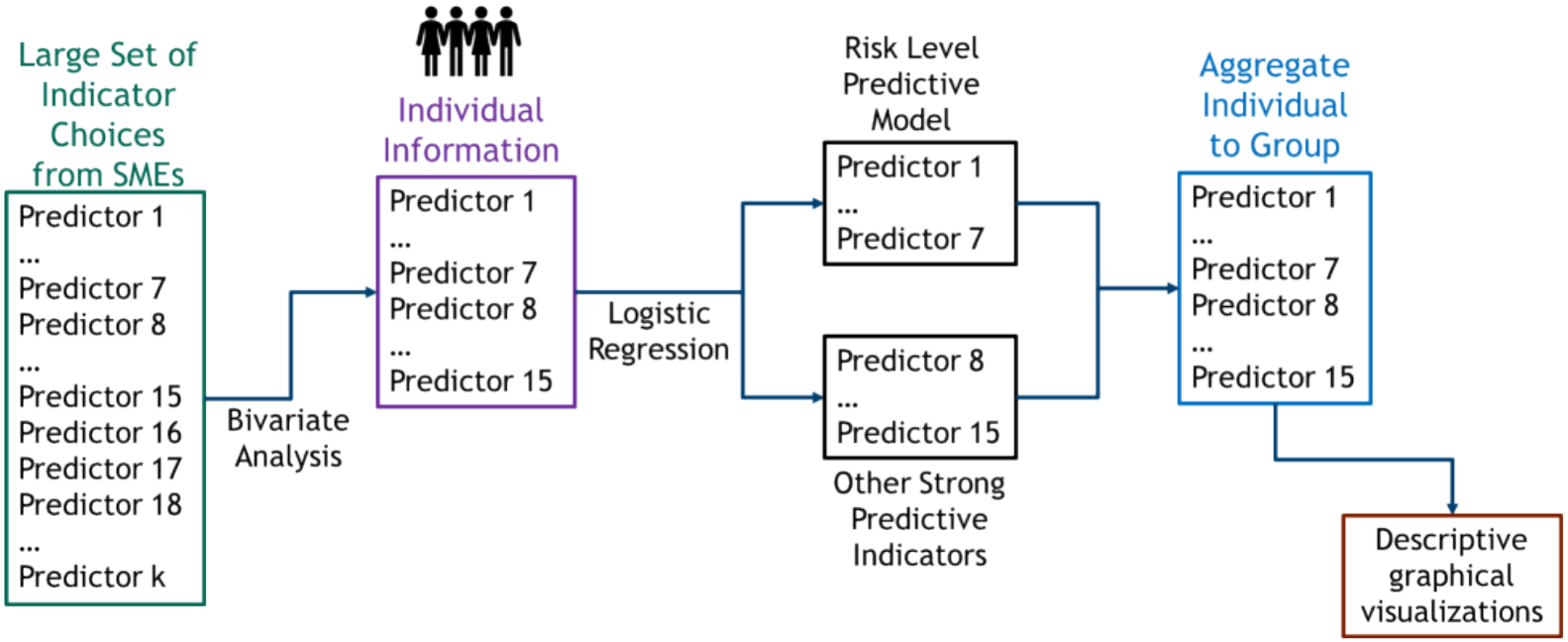
SLIDE TITLE



15 Predictive indicators provide insight toward increased risk

- 7 of the indicators are used by the model to predict risk level.
- Predictions are provided at selected organization level.
- Differentiates which organizations are at highest risk.

SLIDE TITLE

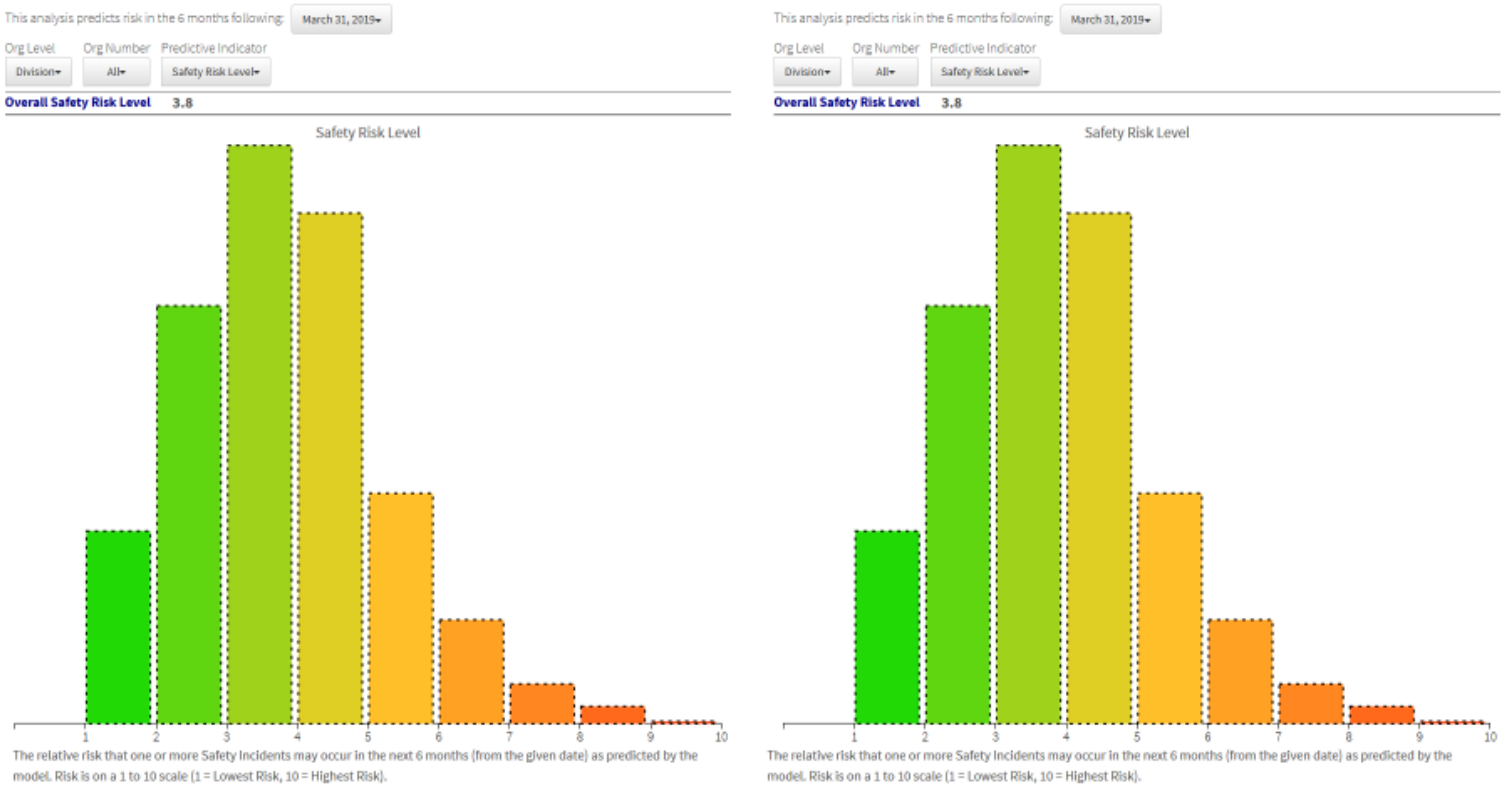


Statistically Validated

- Rigorous validation at time of model development
- Re-validation in 2017
- Originally coded in .NET and SAS macro with incomplete code

Building the Visualization

Beta Version



Built in Microsoft Visual Studio using .NET and SAS macro files

Why Update?

Main goal - help users better understand risk and what is affecting it.

Provide answers to users' typical questions.

- What are the top predictive indicators for risk?
- How are those top predictive indicators changing over time?
- How does one indicator compare to another indicator?

Improve the way users navigate the app by better grouping of relevant information

Enhance the graphs

- Add greater detail (information rich)
- Improve readability with better design choices

Use R!

R is a free software environment for statistical computing and graphics.

R Shiny is an open source R package that provides a web framework for building web applications using R. Shiny helps you build interactive web applications without requiring HTML, CSS, or JavaScript knowledge.

R Shiny Server deploys Shiny applications on the web and behind a firewall.

R Shiny Server Pro makes Shiny applications scalable, password protected and more. This is paid service.

Why Use It?

Free - like free, free

Popular - ranks in 14th place in the TIOBE Index

Flexible - CRAN has more than 14,000 packages available for download

Tidy - Tidyverse is an opinionated collection of R packages designed for data science

Graphics - produce publication-quality graphs

Shiny – build and test web apps without learning JavaScript or HTML

RStudio IDE – a free development environment for R

Support – Lots of free learning material

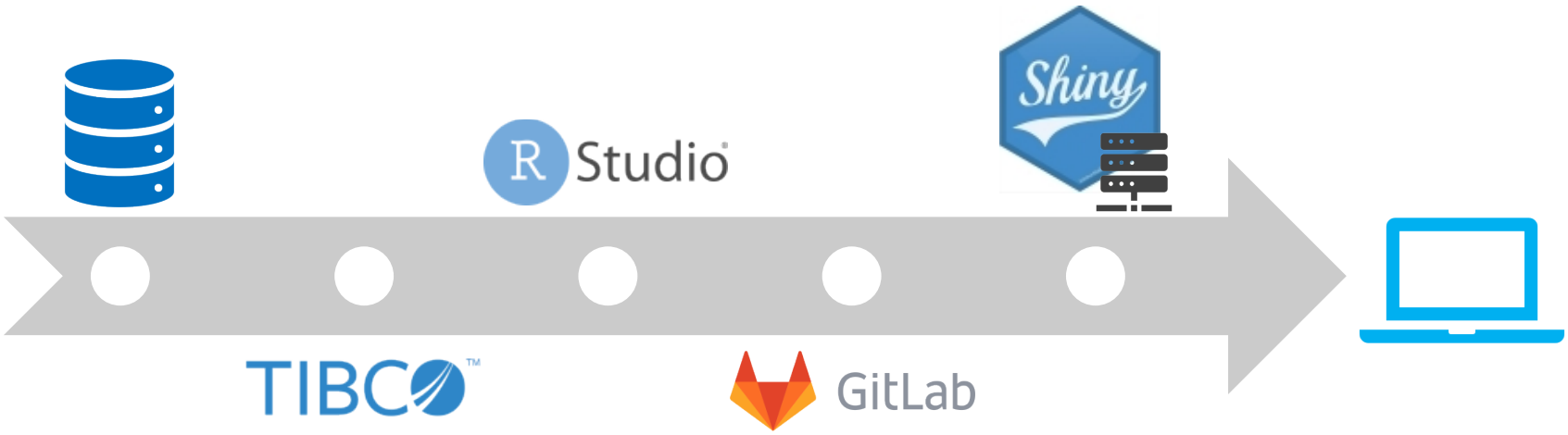
A Single Solution

R for model building

R Shiny for developing viz

R Shiny Server Pro for deploying web app

Architecture



User Centered Design

Requested input from user experience researcher

Conducted Usability Testing

- Round 1: 6 representative users
- Implemented feedback from testing
- Round 2: 3 more representative users
- Implemented feedback from testing
- Deployed final version

Demo

Example

Safety conscious Rene (Org 10774) wants to know what they can do to reduce the number of safety incident in their organization.

1. Signs up for the Assurance for Safety training course.
2. Complete the course.
3. Logs into the Assurance for Safety web app from their desktop.
4. Drills down to their Division 10000, Center 10700, and Group 10770. For legal reasons, Rene cannot view data for their specific organization.
5. See's what the top six predictive indicators are and how they have changed over time.

NOTE: Data and the names of the predictive indicators have been obfuscated.

Filter Options

Show Risk in the 6 Months Following:

March 31, 2019

Select Percentile Rank of Risk Scores:

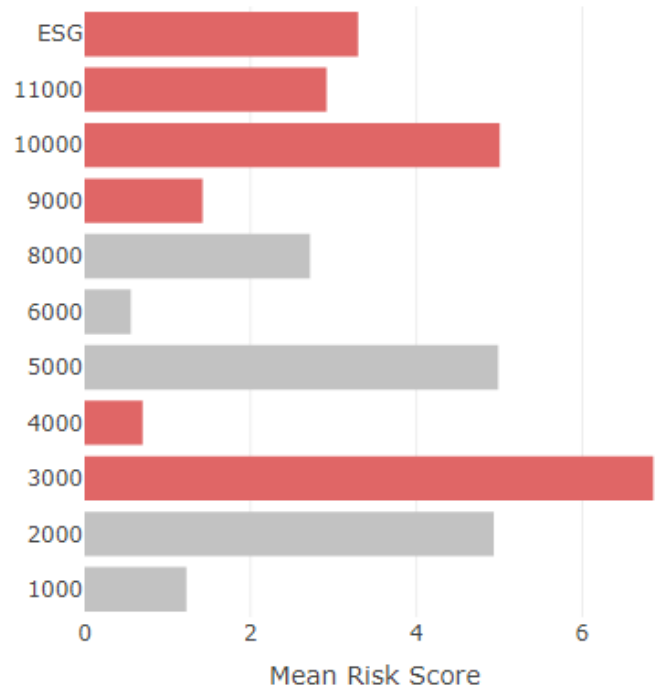
0 50 100

0 10 20 30 40 50 60 70 80 90 100

Number of Preditive Indicators to Show:

3

Division Risk Plot



Division Top Ranked Predictive Indicators.

Click on a Division's bar to display information here.

Center Risk Plot

Center Top Ranked Predictive Indicators

~/GitHub/shiny-assure-safety - Shiny

http://127.0.0.1:4615 | Open in Browser | Publish

Filter Options

Show Risk in the 6 Months Following:

December 31, 2018

Select Percentile Rank of Risk Scores:

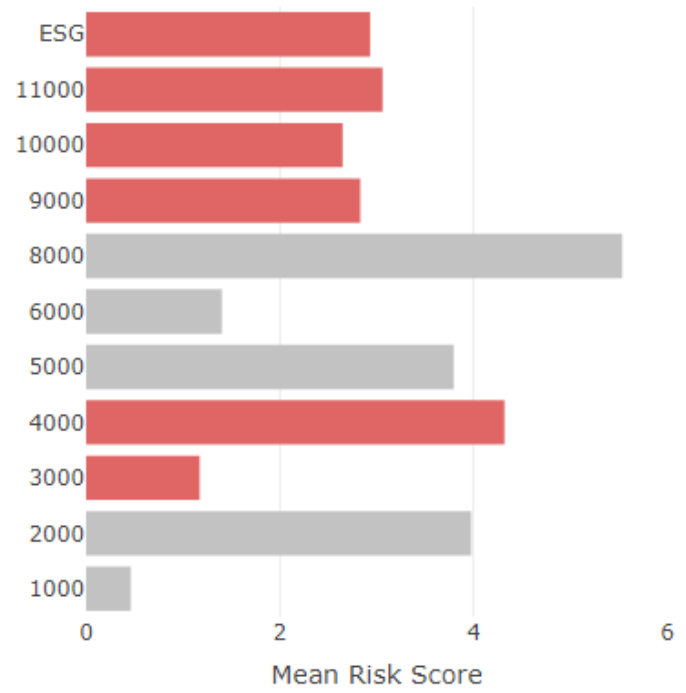
0 50 100

0 10 20 30 40 50 60 70 80 90 100

Number of Preditive Indicators to Show:

6

Division Risk Plot



Division Top Ranked Predictive Indicators.

Click on a Division's bar to display information here.

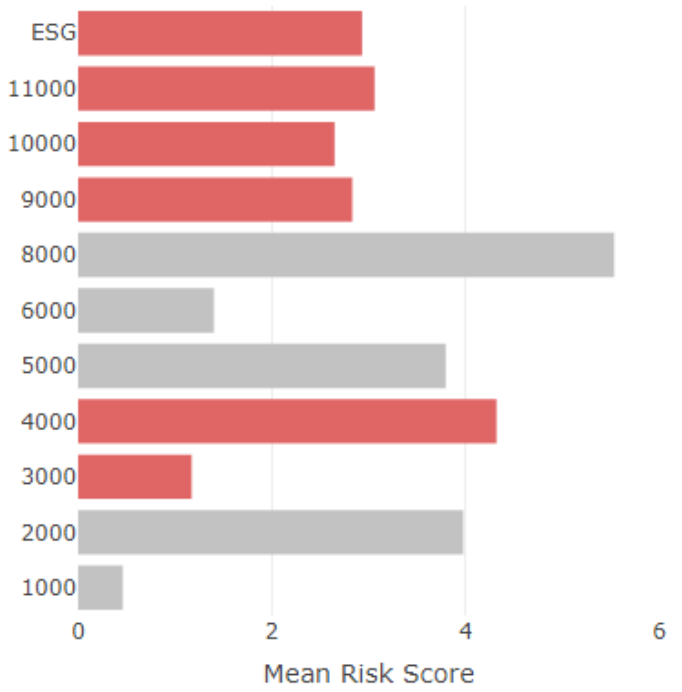
Center Risk Plot

Center Top Ranked Predictive Indicators

~/GitHub/shiny-assure-safety - Shiny

http://127.0.0.1:4615 | Open in Browser | Publish

Division Risk Plot



Division 10000's Top Ranked Predictive Indicators

Rank	Predictive Indicator	Description
1	Time Spent on YouTube	% of employees with <2 years on YouTube is greater than all Sandia
2	Manager's Highest Score in Candy Crush	% of employees whose manager has <2000 high score in Candy Crush is greater than all Sandia
3	Manager's Total Time Playing Candy Crush	% of employees whose manager has <2 years playing Candy Crush is greater than all Sandia
4	Number of Chairs Moving Out of the Organization in the Past 12 Months	% of employees in an organization with 9+ chairs moving out greater than all Sandia
5	Number of Chairs Moving Into the Organization in the Past 12 Months	% of employees in an organization with 9+ chairs moving in greater than all Sandia
6	Number of Retweets in Past 12 Months	% of employees with 1+ retweets in the past 12 months is greater than all Sandia.

View Ranked Histograms | View Time Series Plots

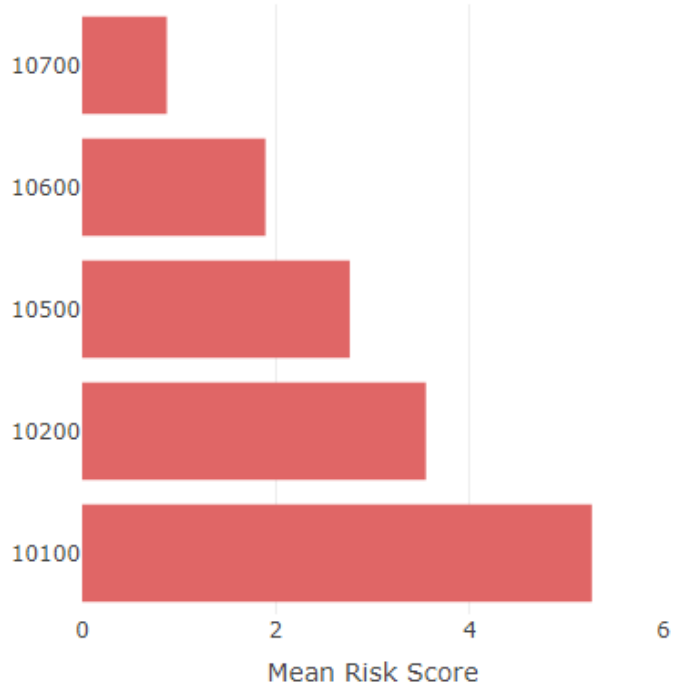
Risk Plot for Centers in Division 10000

Center Top Ranked Predictive Indicators

Click on a Center's bar to display information here.

View Ranked Histograms | View Time Series Plots

Risk Plot for Centers in Division 10000



Center Top Ranked Predictive Indicators

Click on a Center's bar to display information here.

Group Risk Plot

Click on a Group's bar to display its Groups' risk.

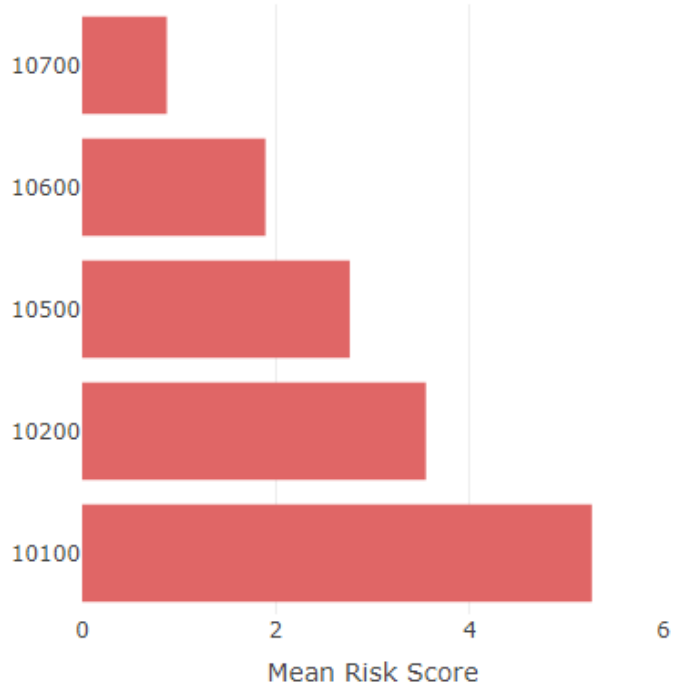
Group Top Ranked Predictive Indicators

Click on a Group's bar to display information here.

ASSURE for Safety (Version 1.2)

View Ranked Histograms | View Time Series Plots

Risk Plot for Centers in Division 10000



Center 10700's Top Ranked Predictive Indicators

Rank	Predictive Indicator	Description
1	Time Spent on YouTube	% of employees with <2 years on YouTube is greater than all Sandia
2	Manager's Highest Score in Candy Crush	% of employees whose manager has <2000 high score in Candy Crush is greater than all Sandia
3	Manager's Total Time Playing Candy Crush	% of employees whose manager has <2 years playing Candy Crush is greater than all Sandia
4	Number of Chairs Moving Out of the Organization in the Past 12 Months	% of employees in an organization with 9+ chairs moving out greater than all Sandia
5	Number of Chairs Moving Into the Organization in the Past 12 Months	% of employees in an organization with 9+ chairs moving in greater than all Sandia
6	Number of Retweets in Past 12 Months	% of employees with 1+ retweets in the past 12 months is greater than all Sandia.

View Ranked Histograms | View Time Series Plots

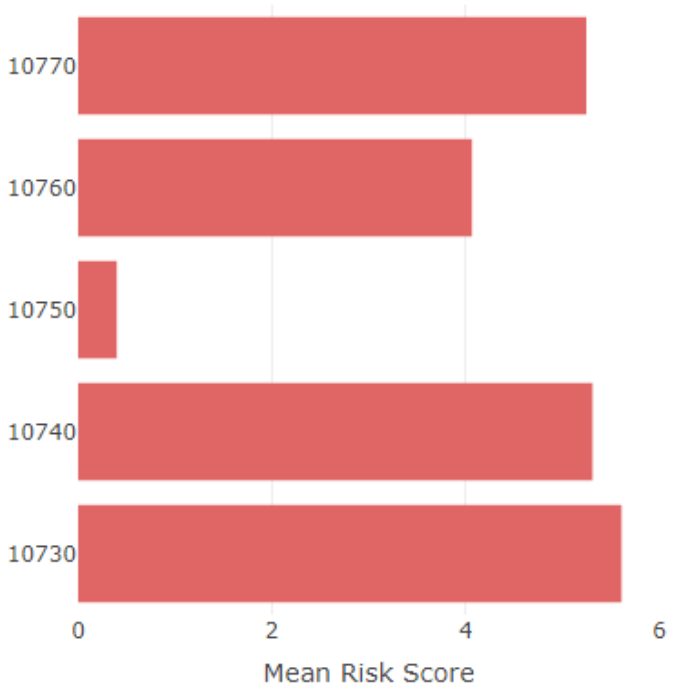
Risk Plot for Groups in Center 10700

Group Top Ranked Predictive Indicators

6	Number of Retweets in Past 12 Months	% of employees with 1+ retweets in the past 12 months is greater than all Sandia.

View Ranked Histograms | View Time Series Plots

Risk Plot for Groups in Center 10700

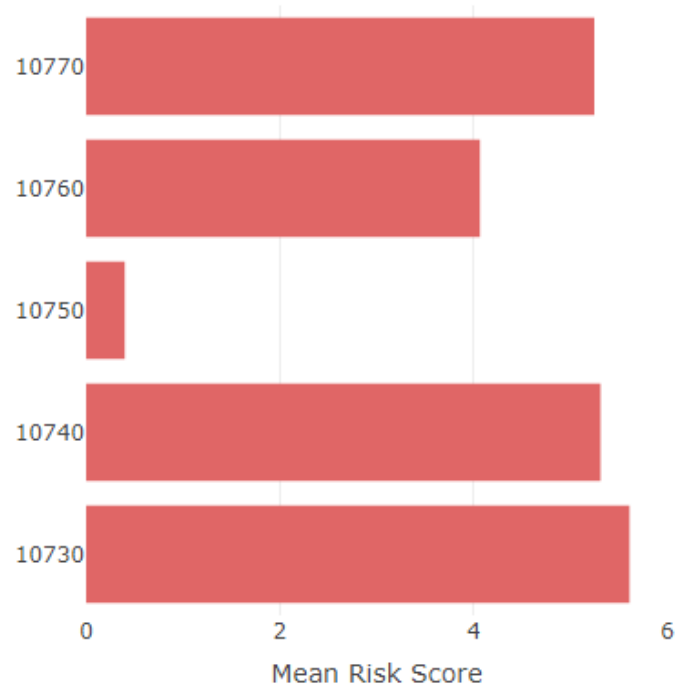


Group Top Ranked Predictive Indicators

Click on a Group's bar to display information here.

ASSURE for Safety (Version 1.2)

Risk Plot for Groups in Center 10700



Group 10770's Top Ranked Predictive Indicators

Rank	Predictive Indicator	Description
1	Manager's Highest Score in Candy Crush	% of employees whose manager has <2000 high score in Candy Crush is greater than all Sandia
2	Time Spent on YouTube	% of employees with <2 years on YouTube is greater than all Sandia
3	Manager's Total Time Playing Candy Crush	% of employees whose manager has <2 years playing Candy Crush is greater than all Sandia
4	Number of Retweets in Past 12 Months	% of employees with 1+ retweets in the past 12 months is greater than all Sandia.
5	Number of Chairs Moving Into the Organization in the Past 12 Months	% of employees in an organization with 9+ chairs moving in greater than all Sandia
6	Number of Bananas Eaten in the Past 48 Months	% of employees with 3+ bananas eaten greater than all Sandia

View Ranked Histograms | View Time Series Plots

ASSURE for Safety (Version 1.2)

~/GitHub/shiny-assure-safety - Shiny

http://127.0.0.1:4615 | Open in Browser | Publish

ASSURE for Safety

≡

Share This Result

Help

How to Use This Tab

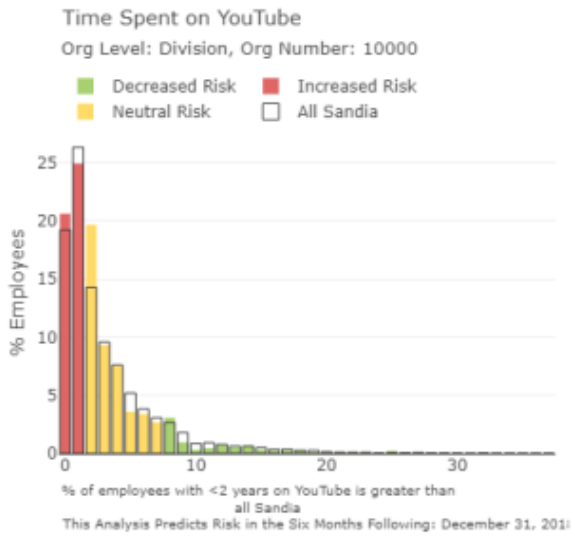
For the specified organization analyzed at the chosen date, the bar plots show the proportion of employees at risk for each value of the selected Predictive Indicator. The x-axis shows the values , and the coloring of the bars shows "zones". Green bars mean that employees in that range of the x-axis are predicted to have a decreased risk, yellow bars indicate neutral impact to risk, while red bars indicate an increased risk.

Return to the Plot Settings Tab to changes these plots.

Plot Legend +

Division

Rank 1 Predictive Indicator

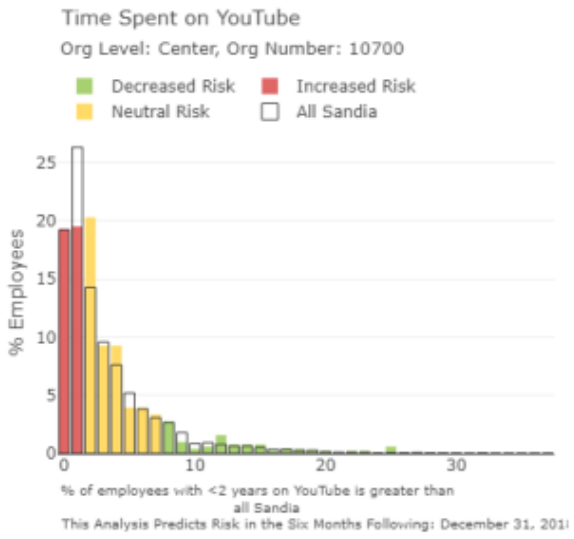


Rank 2 Predictive Indicator

Manager's Highest Score in Candy Crush

Center

Rank 1 Predictive Indicator

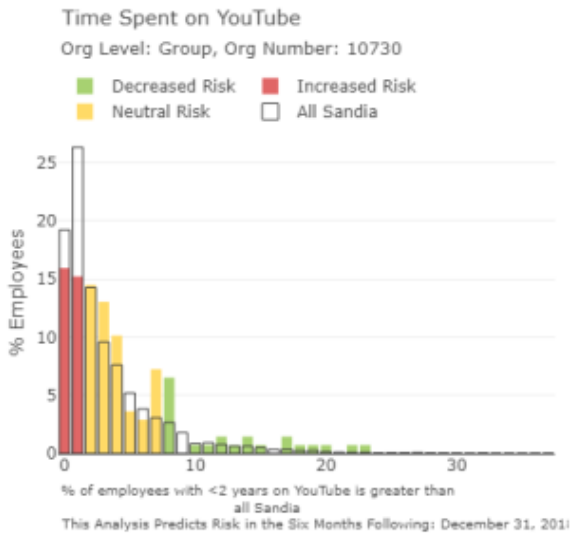


Rank 2 Predictive Indicator

Manager's Highest Score in Candy Crush

Group

Rank 1 Predictive Indicator

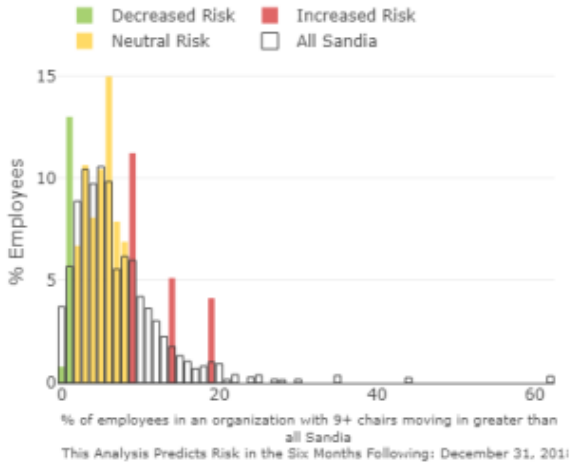


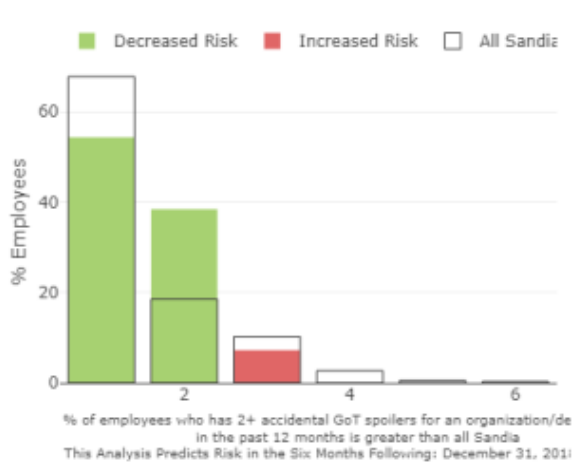
Rank 2 Predictive Indicator

Manager's Total Time Playing Candy Crush

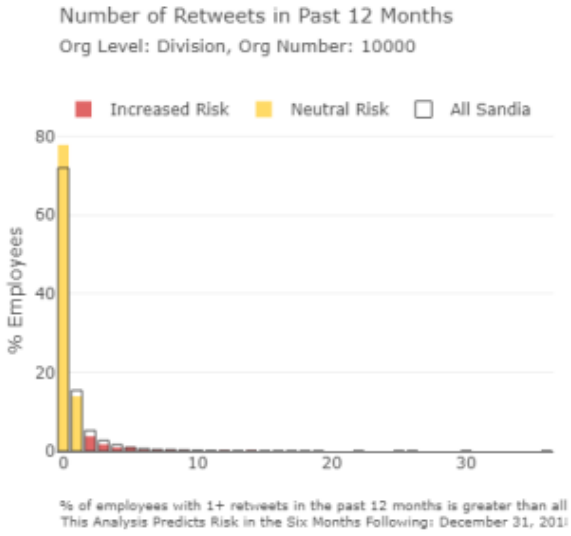
~/GitHub/shiny-assure-safety - Shiny

http://127.0.0.1:4615 | Open in Browser | Publish

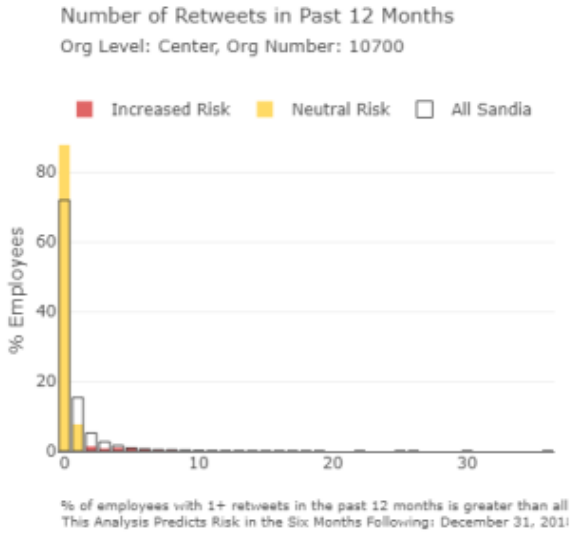




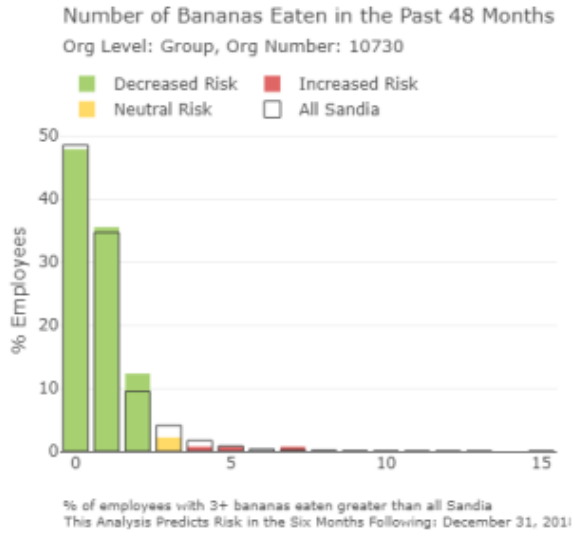
Rank 6 Predictive Indicator



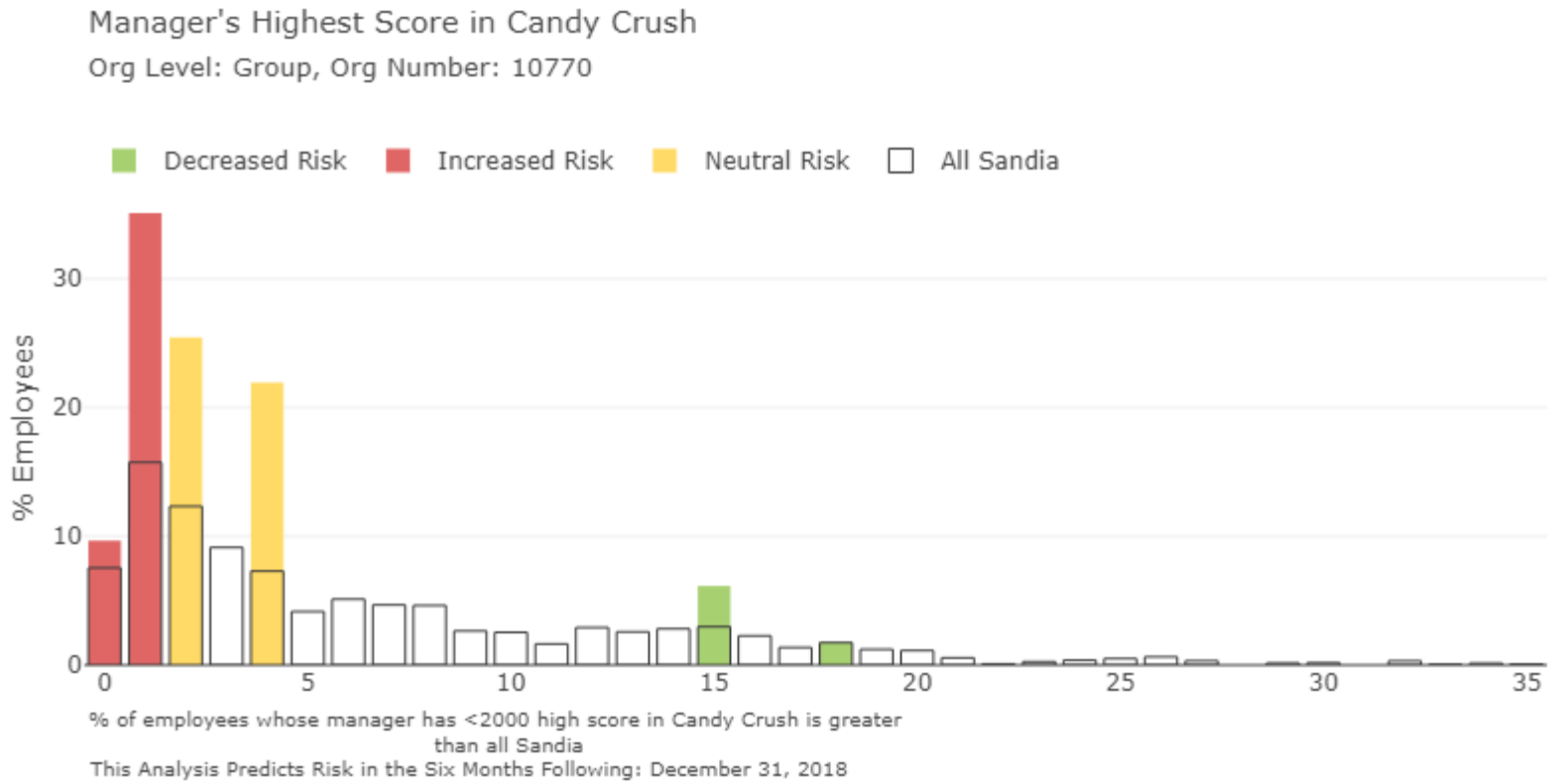
Rank 6 Predictive Indicator



Rank 6 Predictive Indicator



ASSURE for Safety (Version 1.2)

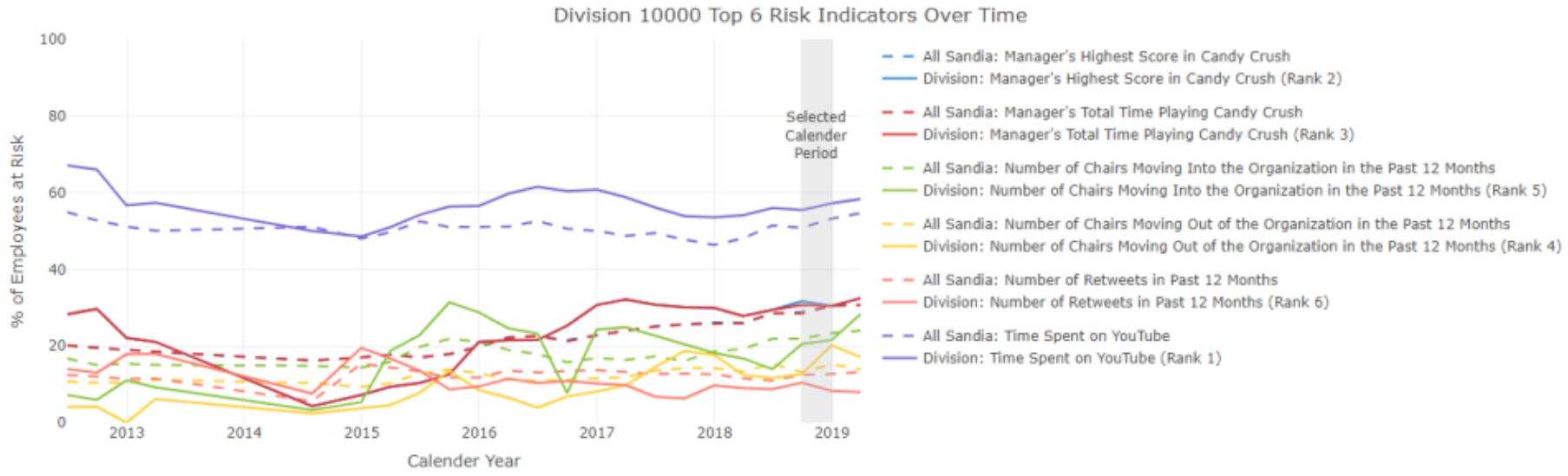

Manager's Highest Score in Candy Crush
Org Level: Group, Org Number: 10770
Decreased Risk
Increased Risk
Neutral Risk
All Sandia
% Employees
30
20
10
0
0
5
10
15
20
25
30
35
% of employees whose manager has <2000 high score in Candy Crush is greater than all Sandia
This Analysis Predicts Risk in the Six Months Following: December 31, 2018

~/GitHub/shiny-assure-safety - Shiny

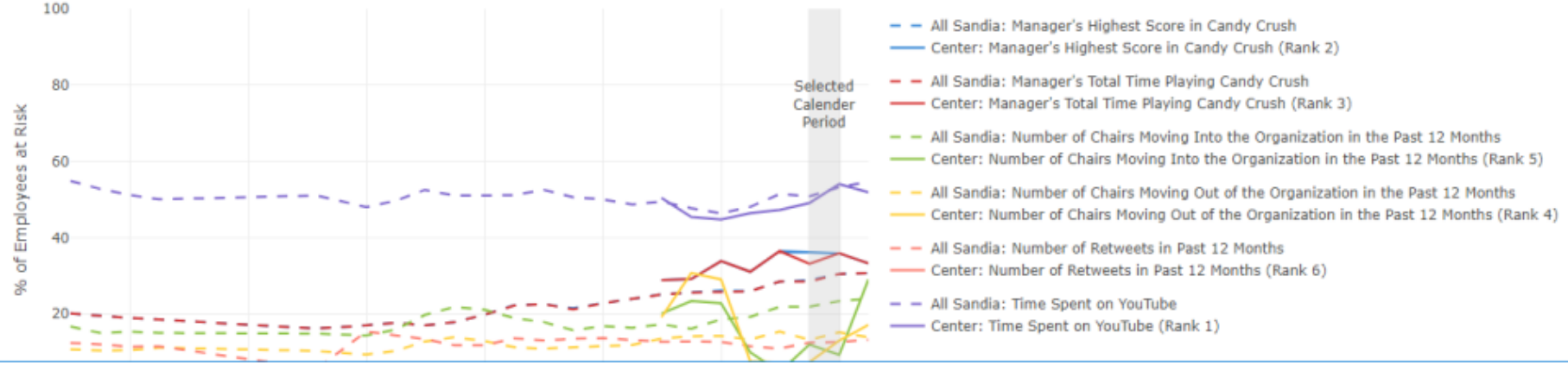
http://127.0.0.1:4615 | Open in Browser | Publish

Return to the Plot Settings Tab to changes these plots.

Division Plot



Center Plot



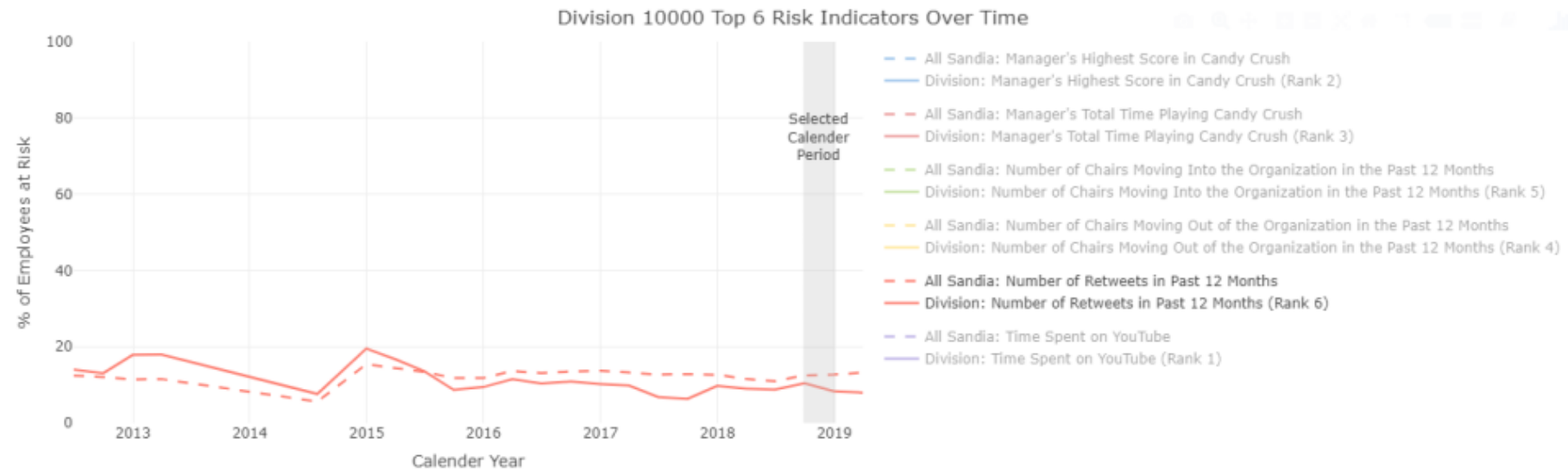
~/GitHub/shiny-assure-safety - Shiny

http://127.0.0.1:4615 Open in Browser

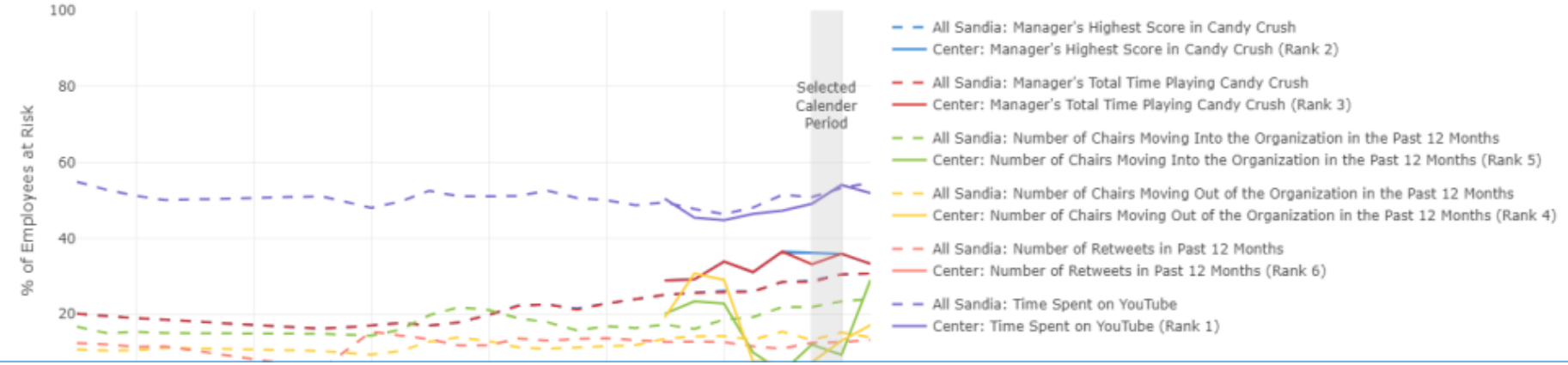
Publish

Return to the Plot Settings Tab to changes these plots.

Division Plot



Center Plot



~/GitHub/shiny-assure-safety - Shiny

http://127.0.0.1:4615 | Open in Browser | Publish

The Risk Detail Comparison tab in the navigation menu will allow users to compare up to two indicators by division, center and group at the same time.

Filter Options

Show Risk in the 6 Months Following: March 31, 2019

Select Org Level: Group

Select Org Number: 10770

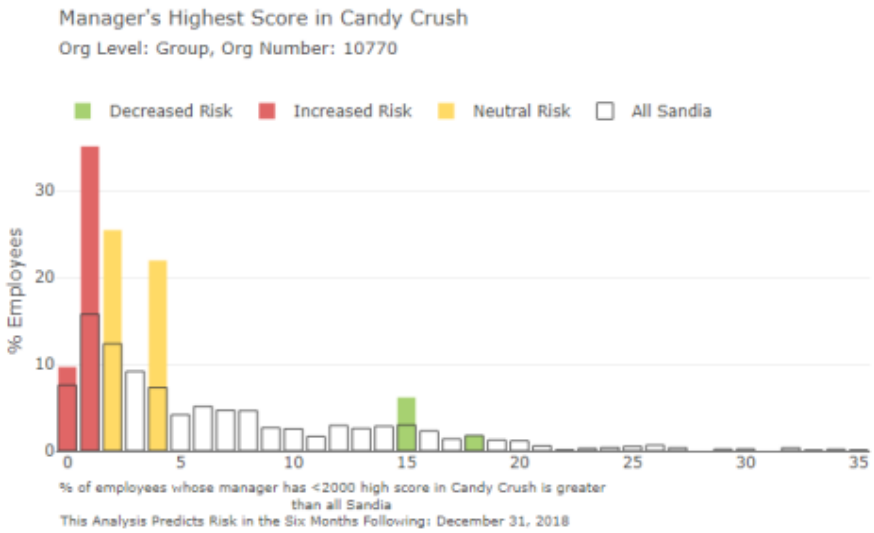
Select Predictive Indicator: Manager's Highest Score in Candy Crush

Select a Plotting Area: Right

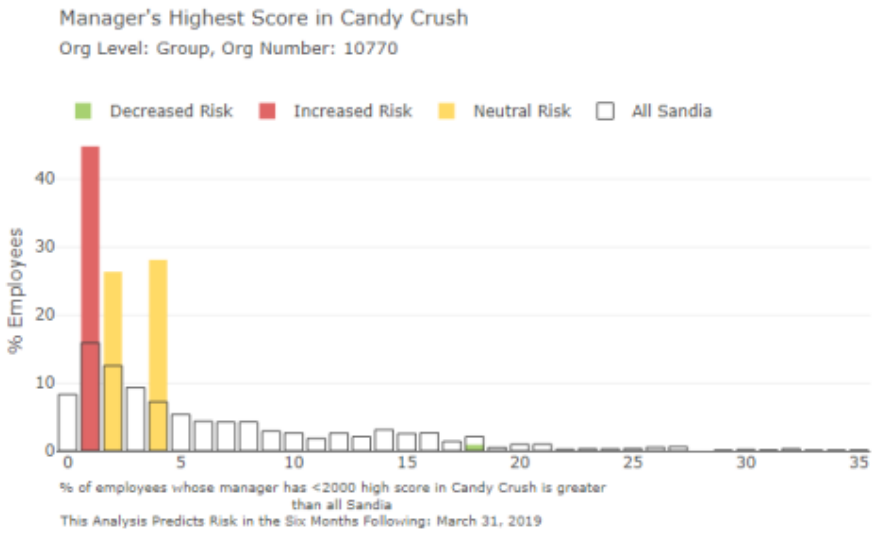
▶ Generate Plot

Plot Legend +

Left Plotting Area



Right Plotting Area



ASSURE for Safety (Version 1.2)

Summary

We wanted to…

- **Inform** – Find organizations at high-risk for one or more safety incidents.
- **Anticipate** – Identify predictive indicators that demonstrate significant relationships to future safety incidents.
- **Be Proactive** – Develop and deploy a visualization (web app) that is user-centered, informative and accessible to drive focus on improvement areas.
- **Prevent** – leading indicators that monitor risk, allowing all to be safety conscious like Rene!

What's Next?

Reevaluate current and new predictive indicators

Try different modeling techniques

What new ways can we use to look at how predictive indicators change over time?

Apply what we learned to the Assurance for Security Model


Questions?