Predicting platelet usage

R in Medicine

Rob Tibshirani
Departments of Biomedical Data Science & Statistics
Stanford University

How many units of platelets will the Stanford Hospital need tomorrow?

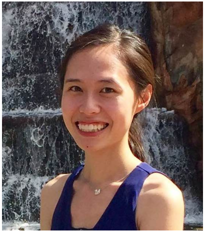
WE WANT YOUR GOLD.

The stuff in your blood, not your bank.

Allison Zemek

Tho Pham

Saurabh Gombar



Leying Guan

Xiaoying Tian

PNAS

Big data modeling to predict platelet usage and minimize wastage in a tertiary care system

Leying Guan[a,1], Xiaoying Tian[a,1], Saurabh Gombar[b], Allison J. Zemek[b], Gomathi Krishnan[c], Robert Scott[d], Balasubramanian Narasimhan[a], Robert J. Tibshirani[a,e,2], and Tho D. Pham[b,d,f,2]

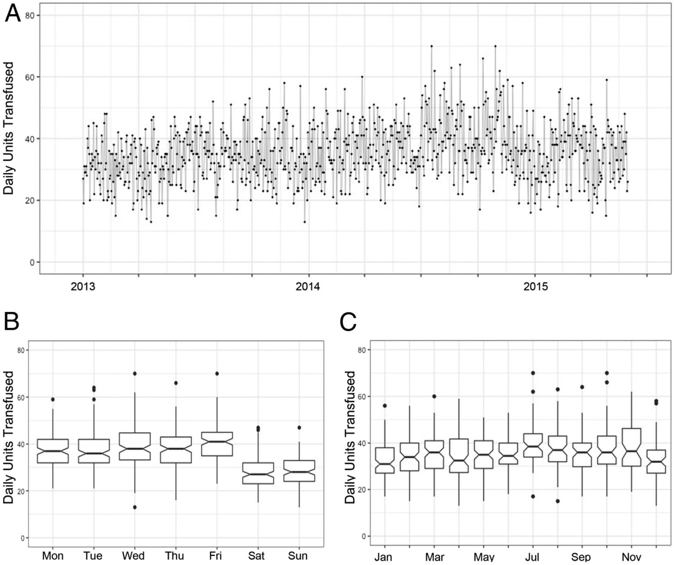
[a]Department of Statistics, Stanford University, Stanford, CA 94305; [b]Department of Pathology, Stanford University, Stanford, CA 94305; [c]Stanford for Clinical Informatics, Stanford University, Stanford, CA 94305; [d]Stanford Hospital Transfusion Service, Stanford Medicine, Stanford, CA 94305; [e]Department of Biomedical Data Science, Stanford University, Stanford, CA 94305; and [f]Stanford Blood Center, Stanford Medicine, Stanford, CA

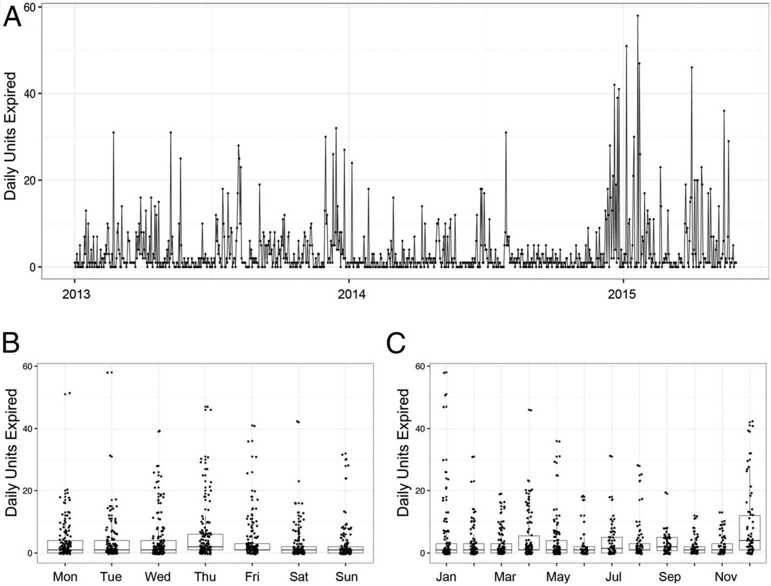
Contributed by Robert J. Tibshirani, August 10, 2017 (sent for review June 25, 2017; reviewed by James Burner, Pearl Toy, and Minh-Ha Tran)

The current system

- Each day, the Stanford blood center collects some number of units (bags) of platelets, based on the estimated needs at Stanford Hospital. The daily needs are estimated "manually".
- Platelets have a **5 day shelf life**, and are safety-tested for 2 days. So they are **usable for just 3 days**, and are discarded after that time.
- Currently about **1400 units (bags) are wasted each year.** That's about 8% of the total number ordered.
- There's rarely any shortage (shortage is bad but not catastrophic)
- *Can we use available information about the hospital to do better?*

Data overview




A
Daily Units Expired
0
20
40
60
2013
2014
2015
B
Daily Units Expired
Mon
Tue
Wed
Thu
Fri
Sat
Sun
C
Daily Units Expired
Jan
Mar
May
Jul
Sep
Nov

Data description

Daily platelet use from 2/8/2013 - 2/8/2015.

- Response: number of platelet transfusions on a given day.
- Covariates:
  1. **Complete blood count (CBC) data**: Platelet count, White blood cell count, Red blood cell count, Hemoglobin concentration, number of lymphocytes, ...
  2. **Census data**: location of the patient, admission date, discharge date, ...
  3. **Surgery schedule data**: scheduled surgery date, type of surgical services, ...
  4. ...

Data description

We first tried to work on the individual patient level, but there were many complications:

1. **Complete blood count (CBC) data**:
   - o 30% of patients have no CBC measurement at all
   - o After being measured, a patient can (1) have a transfusion right away; (2) leave the hospital; (3) come back later in the future but we do not know when.
2. **Census data**:
   - o Often there was no matching medical record number.
3. **Surgery schedule data**:
   - o Often does not match previous data file at the personalized level.
   - o Large percentage of missingness.

Conclusion: **Use aggregated features.**

Feature Construction

- CBC measurement: for each day i and feature j, count the number of patients below the first quartile of the population. Use the average of the past week(11 features).
- CENSUS record: for each day i, count the total number of patients at a location j in the hospital (26 features).

Feature Construction— continued

- PLT transfusion record: for each day i, let y_i be the total number of PLT used at day i. Use the average of past week $\bar{y}_i$ at day i when making prediction(1 feature).
- SURGERY record: for each day i, and count the number of scheduled surgeries at day $i+k$ when making prediction for day future k days, $k = 1, 2, 3$ (17 features).
- Day of the week information: Monday,...,Sunday

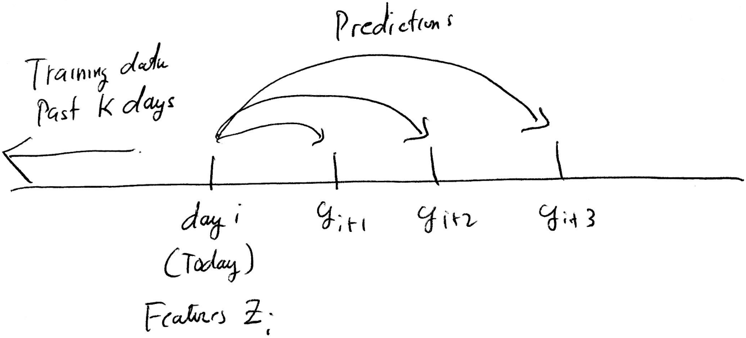
$\Rightarrow$ 61 features in total.

Notation

y_i : actual PLT usage in day i.
x_i : amount of new PLT that arrives at day i.
$r_i(k)$: remaining PLT which can be used in the following k days, $k = 1, 2$
w_i : PLT wasted in day i.
s_i : PLT shortage in day i.

- *Overall objective*: waste as little as possible, with little or no shortage

Our first approach: supervised learning



The Lasso for supervised learning

Given features x_{ij} and an outcome measurement y_i,
the **Lasso** is an estimator defined by the following optimization problem:

$$\underset{\beta_0,\beta}{\text{minimize}} \frac{1}{2}\sum_i (y_i - \beta_0 - \sum_j x_{ij}\beta_j)^2 \qquad \text{subject to} \quad \sum |\beta_j| \leq s$$

- Penalty $\implies$ sparsity (feature selection)
- Convex problem (good for computation and theory)
- Our lab has written a open-source R language package called **glmnet** for fitting lasso models. Available on CRAN.

Our first approach

- Build a supervised learning model (via lasso) to predict use y_i for next three days (other methods like random forests or gradient boosting didn't give better accuracy).
- Starting at day 200, train model. Then use it moving forward, retraining model every month
- We tried training on the prior k days of data. $k =$ all data, or 400, or 150 days.
- Use the estimates $\hat{y}_i$ to estimate how many units x_i to order. Add a buffer to predictions to ensure there is no shortage.
  If $t_i = \hat{y}_i + \hat{y}_{i+1} + \hat{y}_{i+2}$, then amount to order is

$$x_{i+3} = t_i - r_i(1) - r_i(2) - x_{i+1} - x_{i+2}$$

- Works quite well- but (1) choice of buffer is trial and error, and (2) doesn't solve the problem directly (why not?)

A More direct approach

This approach minimizes the waste directly:

$$J(\beta) = \sum_{i=1}^{n} w_i + \lambda||\beta||_1 \quad (1)$$

$$\text{where} \quad (2)$$

$$\text{three days}' \text{ total need} \quad t_i = z_i^T \beta, \quad \forall i = 1, 2, .., n \quad (\textbf{linear predictor}) \quad (3)$$

$$\text{number to order}: x_{i+3} = t_i - r_i(1) - r_i(2) - x_{i+1} - x_{i+2} \quad (4)$$

$$\text{waste} \quad w_i = [r_{i-1}(1) - y_i]_+ \quad (5)$$

$$\text{actual remaining} \quad r_i(1) = [r_{i-1}(2) + r_{i-1}(1) - y_i - w_i]_+ \quad (6)$$

$$r_i(2) = [x_i - [y_i + w_i - r_{i-1}(2) - r_{i-1}(1)]_+]_+ \quad (7)$$

$$\text{Constraint}: \text{ fresh bags remaining} \quad r_i(2) \geq c_0 \quad (\textbf{no shortage allowed}) \quad (8)$$

$$(9)$$

Emphasis shifts from **Prediction of platelet usage** to **Prediction of how many units to order**

This can be shown to be a convex problem (LP). We solve it using standard software in R.

Choice of λ

We choose λ via 8-fold block-wise cross-validation: constraints and targets only involve the remaining 7 folds.
We used the objective function:

$$\sum w_i + 50\{i : r_i < 10\}$$

Important features selected

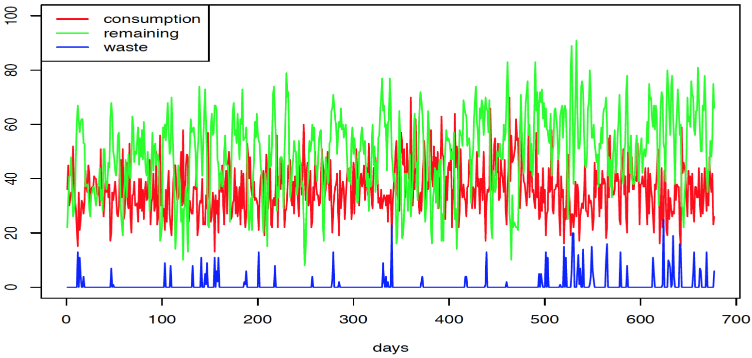
Table: Important features selected

PLT transfusion record $\bar{y}_i$: 5.16	Day of week:Fri -3, Sun +2
HCT: 1.82	RDW:+1
MCHC: -3	RBC:-3
PLT: -3.5	
CENSUS B2: 1	CENSUS C2: +1.5
CENSUS E3: +2	CENSUS H1: +6
CENSUS H2: +2	CENSUS FGR: -1.5

Others:CENSUS E2.ICU, CATH PACU......

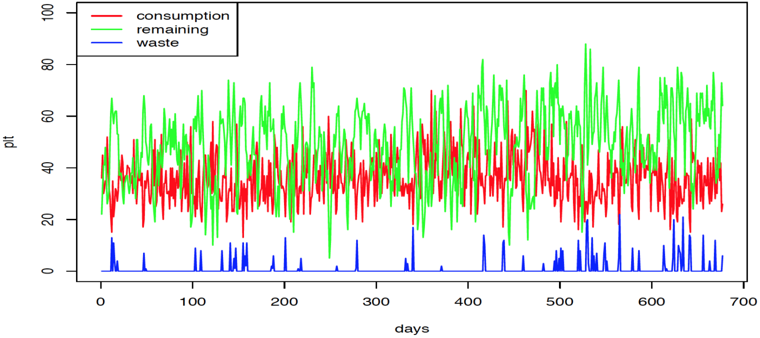
Results: All data

Using all data points from the past as training data: no shortage, waste 389 bags(2.00%) between 2/08/2013 – 2/08/2015



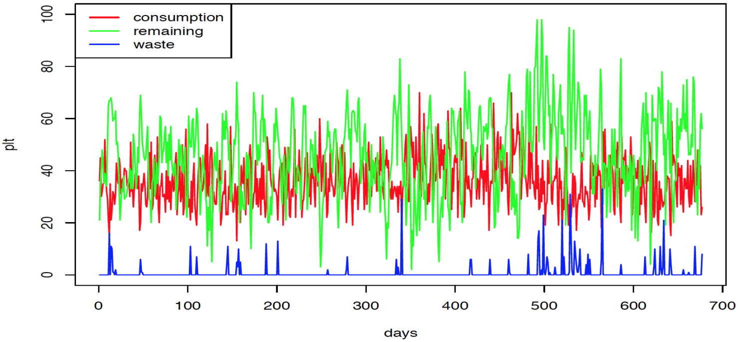
Time window 400

Using 400 data points from the past as training data: no shortage, waste 359 bags(1.85%) between 2/08/2013 – 2/08/2015

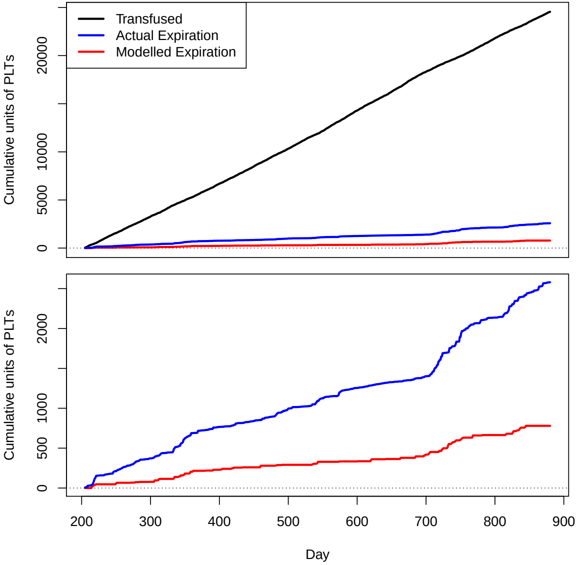


Time window 150

Using 150 data points from the past as training data: no shortage, waste 383 bags(1.97%) between 2/08/2013 – 2/08/2015



Results


Transfused
Actual Expiration
Modelled Expiration
Cumulative units of PLTs
0
5000
10000
20000
Cumulative units of PLTs
0
500
1000
2000
200
300
400
500
600
700
800
900
Day

Summary

- Reducing wastage from 8% to 2% corresponds to a predicted direct savings at Stanford of $350,000/year.
- If implemented nationally could result in approximately $110 million in savings.
- Deployed at Stanford blood center in July 2018
- We are distributing an R package so that blood centers and hospitals can train the model on their own local data , and then deploy the system. Github repo: bnaras/SBCpip
- Application to other time-limited resources?

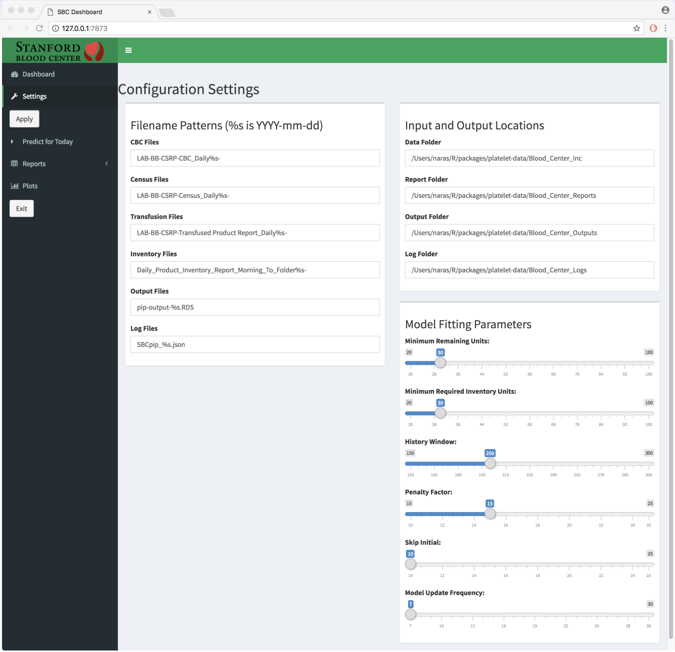
The secrets of our success

1. **Progressive and open-minded medical collaborators** (Tho and Gombur)
2. **Extremely talented and dedicated Statistics graduate students** (Guan and Tian)
3. **First rate statistician/software engineer** (Narashimhan) with great attention to detail

It takes a great team!

R package SBC

- Naras's work. **Important !!** with lots of messy details to tackle
- Consists of a site-specific component for data preparation, and a generic package for model fitting with an R Shiny Interface


SBC Dashboard
127.0.0.1:7873
Stanford Blood Center
Dashboard
Settings
Apply
Predict for Today
Reports
Plots
Exit
Configuration Settings
Filename Patterns (%s is YYYY-mm-dd)
CBC Files
LAB-BB-CSRP-CBC_Daily%s-
Census Files
LAB-BB-CSRP-Census_Daily%s-
Transfusion Files
LAB-BB-CSRP-Transfused Product Report_Daily%s-
Inventory Files
Daily_Product_Inventory_Report_Morning_To_Folder%s-
Output Files
pip-output-%s.RDS
Log Files
SBCpip_%s.json
Input and Output Locations
Data Folder
/Users/naras/R/packages/platelet-data/Blood_Center_inc
Report Folder
/Users/naras/R/packages/platelet-data/Blood_Center_Reports
Output Folder
/Users/naras/R/packages/platelet-data/Blood_Center_Outputs
Log Folder
/Users/naras/R/packages/platelet-data/Blood_Center_Logs
Model Fitting Parameters
Minimum Remaining Units:
20 30 100
Minimum Required Inventory Units:
20 30 100
History Window:
150 200 300
Penalty Factor:
10 15 25
Skip Initial:
10 25
Model Update Frequency:
7 30

SBC Dashboard

127.0.0.1:4393

STANFORD BLOOD CENTER

- Dashboard
- Settings
- Predict for Today

Date

2018-07-09

Predict

- Reports
- Plots

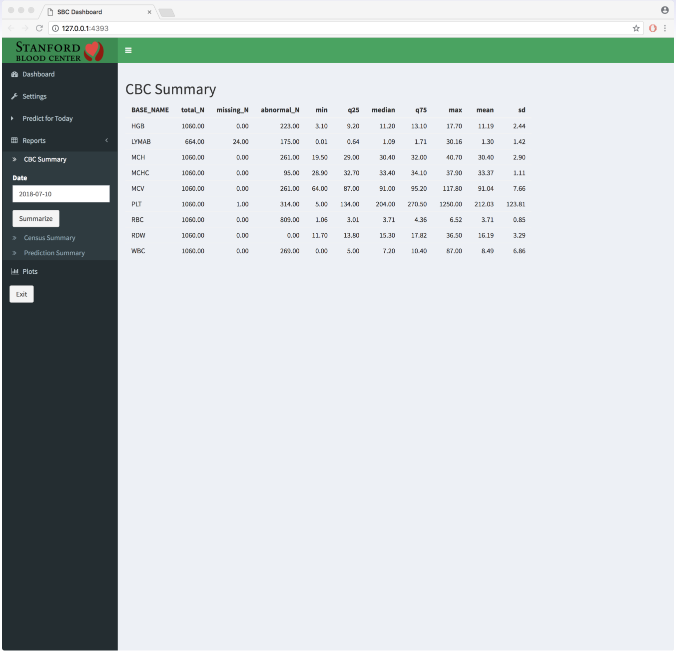
Exit

Prediction Result

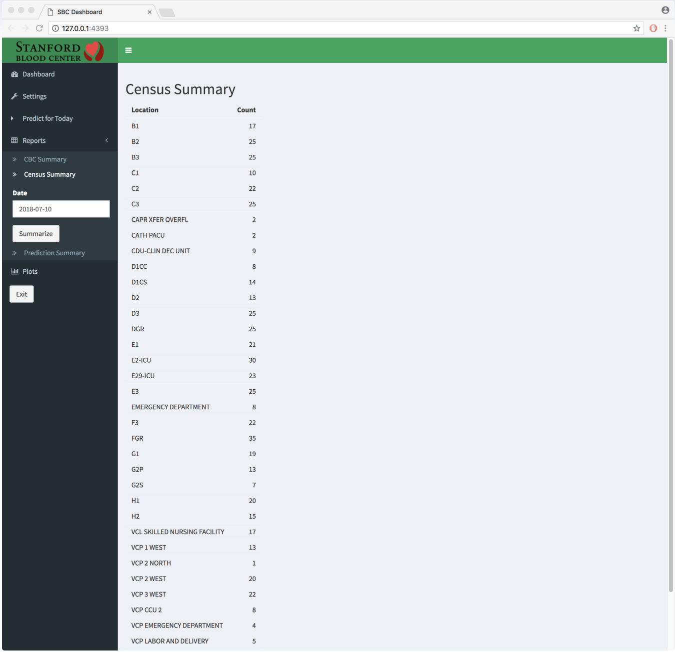
date	t_pred
2018-07-08	193.00

Prediction Log

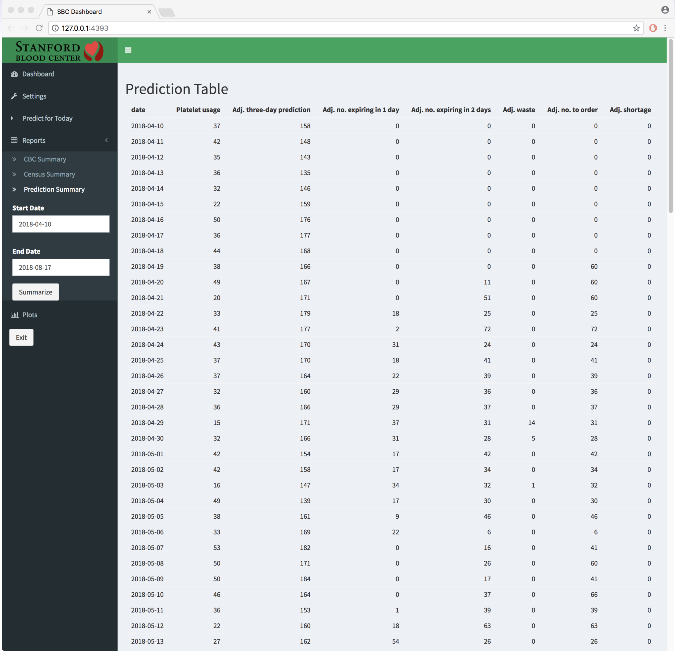
timestamp	log_lvl	log_msg	log_detail
2018-08-17 17:54:03	INFO	Initial log	
2018-08-17 17:54:03	INFO	Step 1. Loading previously processed data on 2018-07-08	
2018-08-17 17:54:04	INFO	Step 2. Processing incremental data for date 2018-07-09	
2018-08-17 17:54:04	INFO	Processing LAB-BB-CSRP-CBC_Daily2018-07-09-08-29-38.csv	
2018-08-17 17:54:04	INFO	Writing Report /Users/naras/R/packages/platelet-data/Blood_Center_Reports/LAB-BB-CSRP-CBC_Daily2018-07-09-08-29-38-summary.xlsx	
2018-08-17 17:54:04	INFO	Processing LAB-BB-CSRP-Census_Daily2018-07-09-08-02-10.csv	
2018-08-17 17:54:04	INFO	Writing Report /Users/naras/R/packages/platelet-data/Blood_Center_Reports/LAB-BB-CSRP-Census_Daily2018-07-09-08-02-10-summary.xlsx	
2018-08-17 17:54:04	INFO	Processing LAB-BB-CSRP-Transfused Product Report_Daily2018-07-09-08-05-44.txt	
2018-08-17 17:54:04	INFO	Writing Report /Users/naras/R/packages/platelet-data/Blood_Center_Reports/LAB-BB-CSRP-Transfused Product Report_Daily2018-07-09-08-05-44-summary.xlsx	
2018-08-17 17:54:04	INFO	Processing Daily_Product_Inventory_Report_Morning_To_Folder2018-07-09-07-30-00.xls for 2018-07-08 23:59:59 inventory	
2018-08-17 17:54:04	INFO	Writing Report /Users/naras/R/packages/platelet-data/Blood_Center_Reports/Daily_Product_Inventory_Report_Morning_To_Folder2018-07-09-07-30-00-summary.xlsx	
2018-08-17 17:54:04	INFO	Step 3. Adding new increment to previous data	
2018-08-17 17:54:04	INFO	Step 3a. Creating CBC features	
2018-08-17 17:54:06	INFO	Step 3b. Creating training/prediction dataset	
2018-08-17 17:54:06	INFO	Step 4. Checking model age	
2018-08-17 17:54:06	INFO	Step 4.1. Using previous model and scaling	
2018-08-17 17:54:06	INFO	Step 5. Predicting and bumping model age	
2018-08-17 17:54:06	INFO	Step 6. Save results for next day	



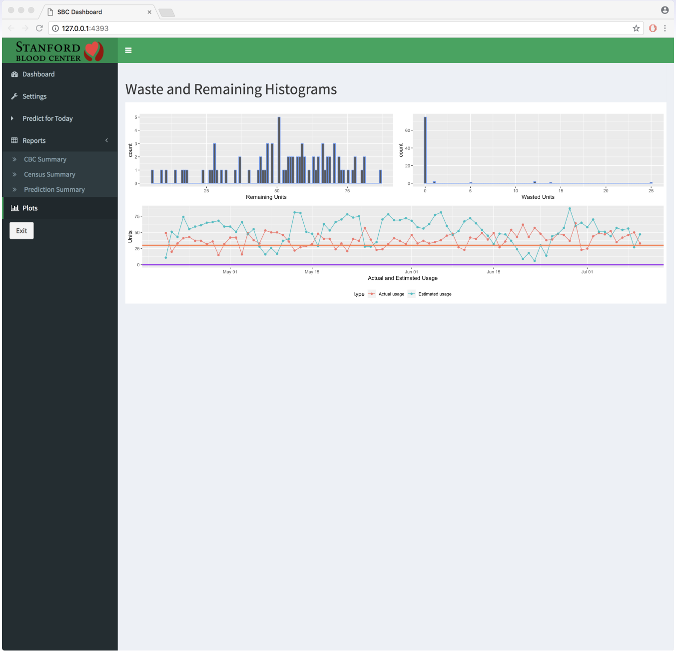
BASE_NAME	total_N	missing_N	abnormal_N	min	q25	median	q75	max	mean	sd
HGB	1060.00	0.00	223.00	3.10	9.20	11.20	13.10	17.70	11.19	2.44
LYMAB	664.00	24.00	175.00	0.01	0.64	1.09	1.71	30.16	1.30	1.42
MCH	1060.00	0.00	261.00	19.50	29.00	30.40	32.00	40.70	30.40	2.90
MCHC	1060.00	0.00	95.00	28.90	32.70	33.40	34.10	37.90	33.37	1.11
MCV	1060.00	0.00	261.00	64.00	87.00	91.00	95.20	117.80	91.04	7.66
PLT	1060.00	1.00	314.00	5.00	134.00	204.00	270.50	1250.00	212.03	123.81
RBC	1060.00	0.00	809.00	1.06	3.01	3.71	4.36	6.52	3.71	0.85
RDW	1060.00	0.00	0.00	11.70	13.80	15.30	17.82	36.50	16.19	3.29
WBC	1060.00	0.00	269.00	0.00	5.00	7.20	10.40	87.00	8.49	6.86



Location	Count
B1	17
B2	25
B3	25
C1	10
C2	22
C3	25
CAPR XFER OVERFL	2
CATH PACU	2
CDU-CLIN DEC UNIT	9
D1CC	8
D1CS	14
D2	13
D3	25
DGR	25
E1	21
E2-ICU	30
E29-ICU	23
E3	25
EMERGENCY DEPARTMENT	8
F3	22
FGR	35
G1	19
G2P	13
G2S	7
H1	20
H2	15
VCL SKILLED NURSING FACILITY	17
VCP 1 WEST	13
VCP 2 NORTH	1
VCP 2 WEST	20
VCP 3 WEST	22
VCP CCU 2	8
VCP EMERGENCY DEPARTMENT	4
VCP LABOR AND DELIVERY	5



date	Platelet usage	Adj. three-day prediction	Adj. no. expiring in 1 day	Adj. no. expiring in 2 days	Adj. waste	Adj. no. to order	Adj. shortage
2018-04-10	37	158	0	0	0	0	0
2018-04-11	42	148	0	0	0	0	0
2018-04-12	35	143	0	0	0	0	0
2018-04-13	36	135	0	0	0	0	0
2018-04-14	32	146	0	0	0	0	0
2018-04-15	22	159	0	0	0	0	0
2018-04-16	50	176	0	0	0	0	0
2018-04-17	36	177	0	0	0	0	0
2018-04-18	44	168	0	0	0	0	0
2018-04-19	38	166	0	0	0	60	0
2018-04-20	49	167	0	11	0	60	0
2018-04-21	20	171	0	51	0	60	0
2018-04-22	33	179	18	25	0	25	0
2018-04-23	41	177	2	72	0	72	0
2018-04-24	43	170	31	24	0	24	0
2018-04-25	37	170	18	41	0	41	0
2018-04-26	37	164	22	39	0	39	0
2018-04-27	32	160	29	36	0	36	0
2018-04-28	36	166	29	37	0	37	0
2018-04-29	15	171	37	31	14	31	0
2018-04-30	32	166	31	28	5	28	0
2018-05-01	42	154	17	42	0	42	0
2018-05-02	42	158	17	34	0	34	0
2018-05-03	16	147	34	32	1	32	0
2018-05-04	49	139	17	30	0	30	0
2018-05-05	38	161	9	46	0	46	0
2018-05-06	33	169	22	6	0	6	0
2018-05-07	53	182	0	16	0	41	0
2018-05-08	50	171	0	26	0	60	0
2018-05-09	50	184	0	17	0	41	0
2018-05-10	46	164	0	37	0	66	0
2018-05-11	36	153	1	39	0	39	0
2018-05-12	22	160	18	63	0	63	0
2018-05-13	27	162	54	26	0	26	0


SBC Dashboard
127.0.0.1:4393
STANFORD BLOOD CENTER
Dashboard
Settings
Predict for Today
Reports
CBC Summary
Census Summary
Prediction Summary
Plots
Exit
Waste and Remaining Histograms
count
Remaining Units
count
Wasted Units
Units
May 01
May 15
Jun 01
Jun 15
Jul 01
Actual and Estimated Usage
type
Actual usage
Estimated usage