What an ANOVA does is let you compare 2 *or more* treatments but without trickery and in one go. Say one group which is sober, another which has just had a couple of glasses of wine, a third group being totally paralytic and, just for fun, a fourth group of gerbils on mind-altering hallucinogenics.  
  
With an ANOVA, you could compare the game scores with a single test. It will tell you whether there are differences in these conditions or not, but not where.

‘Temperature series’

When we want to generate alarms

Going out of range is okay

Defrost cycles in freezer- get above freezing temperature to remove humidity

0 degrees half an hour, expect next defrost cycle don’t send alarm

Mainly in pharma

Makes and models, looking for patterns and consistency, make and model

What modles to use

Clean model?

A priori?

Make and model error

Human input errors

1. Build tools that allow you to **ingest data,** and separate out individual data, different units (refridgerator unist), from plotting the time series, start to come up with classes of units. Difference between units. One big difference is coolers and freezers. What parameter would tell you difference between cooler and freezer. Just some stats test. Ie Mean temperature. **Will be go to tool**, **go back to plotting tool in weird data, some units where number of datapoinst is extremely low**

Two things to look at while looking at dota

1.Reptition in patterns of the data- characterize patterns in data, this unit is no longer operating in the way it used to

2. Can we look at aspects of the unit relative to the make and model.

#plot histograms first to test normal distribution

#40 to -20 that is not normal, large deviation hard to break 3sd, devices could have large temperature

#swings

#missing another category of sensor/reading/problems

For df1..10 regular cooling cycle vs short term cooling. Longer term *defrost* (condensation clearing on coils) How to get rid of this noise? NOT normal. SD dominated by cooling cycle extremes

Histogram just check for normality/Gaussian

Go back and pick highest SD, relative to low sd see differences

Week by week, what’s the mean and standard deviation. Bias

Uses cases for bias

iDegrees 35-45 F maintain- defrost cycle and cooling cycle may throw mean to required temperature for medicatoin

iiCooler to freezer sensor change

iiiSensor turned on but not on unit (benign)

iv Cooler is unreliable, three degrees range, but mean is constantly changing

time/hangout

Device changing operating characteristic once a week

If perfectly normal- don’t use

Alright- then use 3sd

cyclical\*

acceptable threshold for each sensor

identify outliers, how many cycles do you need to see, cooling period

-3/24

I'll be giving you a set of advices in terms of how to program this in R, that could be useful for you later on:   
1. Used data.table instead of dplyr or plyr. It's faster for everything, is practically the only good thing in R for data manipulation that is better than pandas (python).  
2. Use fread to read the files, it's faster.   
library(data.table)  
?fread  
3. Once you read the files and rbind them, save them as an RData file. This will make you loose less time at re-running the code. You'll iterate much faster.

4. With data.table you don't need to compute an aggregation and then merge it to the large dataset. You can do this much more efficiently.   
  
The := symbol creates a new column (by reference). With the mean of the original reading by SensorID.   
  
DT[, m := mean(OriginalReading), by=SensorID]  
  
data.table works very similar to SQL, the main idea is:  
DT[ WHERE , SELECT, GROUP BY][HAVING]  
  
I think in your case, you need to think more of the time series rather than just computing averages.   
I would first make some plots of the time serie to understand the problem better. The next step would be to define what is an anomaly, for each case. Then think of models or perhaps somehting simpler to detect them  
There are 29794 observations without the complete date, any ideas what happened there?

I've read the data and saved it as an RData file. You should have RStudio installed.   
- Then the data\_ingestion.R script has the relevant parts of your original script.   
- The analysis.R script has code to make the plots and the anomaly detection analysis.   
- R\_libraries.R has all the packages needed. I think all of them are on cran. install them with   
  
Create a vector with the packages name and use the install.packages function.  
pkgs = c('data.table', 'ggplot2', ...)  
install.packages(pkgs)

Series can be ‘stationary’ in terms of mean (going up and down), or variance

What does variance tell us about sensor unit etc?

Python?  
Where do you find relevant packages?

Decompose seasonality (2 days ago, 2 hours ago, 2 months ago, any repeating period- weekly seasonality ie people buying things more often on Mondays) , trend, remainder (whatever is not explained by two other components) gaussian noise

#zoo for time series

Standard time series procedure called SDL

Trying different sensors  
Google: Ideally looking at anamoly detection in remainder in time series, twitter anamoly package in R

Your time series are pretty damn consistent, unlike economic trends, so your trend line doesn’t really tell you much about anamolies

<https://blog.twitter.com/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series>

twitter example- underlying stationarity has changed, although variance might not have

How long can a fridge RETAIN its temperature bounds, tight sd of 2,

Underlying model has changed, continuity has changed