Analyzing Heat and Humidity at LA International Airport

Makaylah Cowan

5/1/2020

Table of Contents



Aircraft Taking Off on Sunny Day Source: WallpaperSafari.

# Introduction

Relative humidity is the percentage of water vapor in the air. For example, really saturated air that cannot hold any more water would have a relative humidity of 100%. Relative humidity is often affected by air temperature. If air temperature increases then the air can hold more water molecules and the relative humidity decreases. Conversely, if air temperature decreases then the air can hold less molecules and relative humidity increases, thus air temperature and humidity are directly related. When thinking of the most ideal flying conditions, typically we think of a bright, sunny, no-wind day, and while this can be the case, it ignores a hidden problem that can arise from these same conditions. Along with sunny days, summer brings heat and humidity which has the potential to cause serious issues with aircraft performance. These issues, called ‘density altitude’, stems from aerodynamics; humidity and hot air can cause the air to be less dense which makes it harder to sustain flight.

In this report, I will check assumptions, look at the correlation between temperature and humdiity, and analyze how the variables have changed individually at Los Angeles international airport across the years.

# Methods

This report was created based on a compilation of data hosted by the National Oceanic and Atmospheric Administration ([NOAA](ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/)). It features hourly meteorological data from LOS ANGELES INTERNATIONAL AIRPORT (LAX), with variables including temperature (Celsius), relative humidity, quantity of time for which precipitation was measured, depth of precipitation, and whether the depth was a trace amount. Additional documentation can be found in this NOAA [PDF](https://www1.ncdc.noaa.gov/pub/data/ish/ish-format-document.pdf) under ‘Precipitation Data’. Along with missing values, values for 1996 were removed as it did not have complete, annual observations. Unless specified as otherwise, the data was faceted by year giving each group n=235.

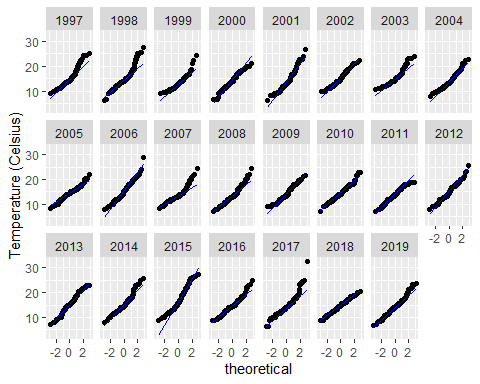
# Results

# Setup  
library(dplyr)  
library(plyr)  
library(ggplot2)  
library(lubridate)  
library(lattice)  
  
Airport <- readRDS("../Data/LA\_airport.rds")  
  
cleandata <- Airport %>%   
 na.omit() %>%   
 filter(!grepl(9, aa1\_3) & !grepl(1, aa1\_3) & !grepl(999.9, aa1\_2) & !grepl(99, aa1\_1), !grepl(9, aa1\_4)) %>%   
 mutate(year = year(time),  
 month = factor(months(time),levels = c("January", "February", "March", "April", "May",   
 "June", "July", "August", "September",  
 "October", "November", "December")))  
  
cleandata <- cleandata[!(cleandata$year==1996),]

### Assumptions

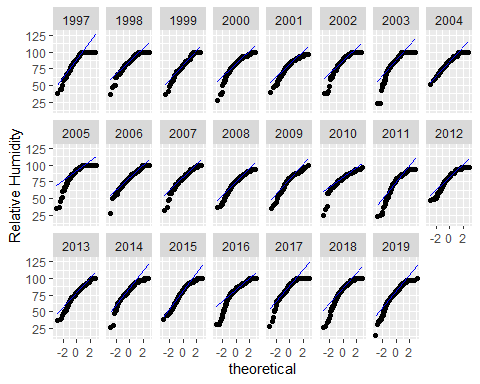
This report uses Anova tests for the temperature and relative humidity variables which assumes sample independence, normality, and equality of variances. The sample independence requirement is met. To check for normality, we will plot the variables against a theoretical quantile-quantile plot.

ggplot(cleandata, aes(sample = temp)) + geom\_qq(distribution = qnorm) +   
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "blue") + ylab("Temperature (Celsius)") +  
 facet\_wrap(~ year, nrow = 3) + ylab("Temperature (Celsius)")



The temperature data appears to generally follow the line and suggesting normality. 1999, 2007, 2015, and 2017 have some endpoints straying which is important to keep in mind as we continue the analysis.

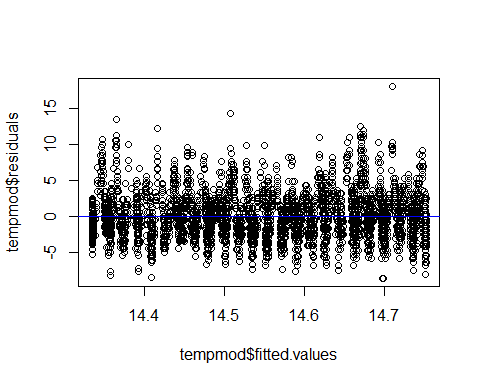
ggplot(cleandata, aes(sample = rh)) + geom\_qq(distribution = qnorm) +   
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "blue") + ylab("Relative Humidity") +  
 facet\_wrap(~ year, nrow = 3) + ylab("Relative Humidity")



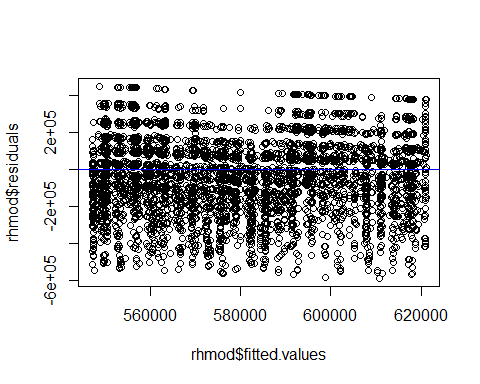
The humidity data shows extreme division from the line. The shape of the endpoints suggest left skewed data; meaning that the median value is greater than the mean value. To fix this we will cube the relative humidity data and use that going forward.

Lastly, we will test for the equality of variances for each variable by generating a residual vs. fitted values plot.

tempmod <- lm(temp~time,cleandata)  
plot(tempmod$fitted.values,tempmod$residuals)  
abline(h=0, col="blue")



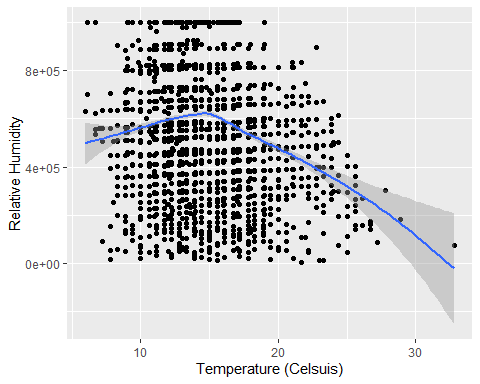
rhmod <- lm(rh^3~time,cleandata)  
plot(rhmod$fitted.values,rhmod$residuals)  
abline(h=0, col="blue")



From these plots, we can see that there is no apparent trend in the residuals for either variable which means that there is homogeneity in variance, so this assumption is met.

### Correlation

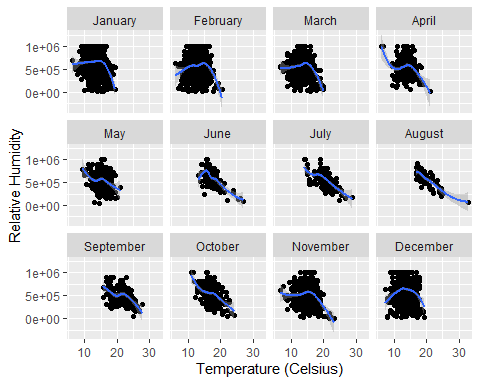
ggplot(cleandata, aes(x = temp, y = rh^3)) + geom\_point() + geom\_smooth(method = "loess") + xlab("Temperature (Celsuis)") +  
 ylab("Relative Humidity")



As expected, the relationship between temperature and humidity is direct and negative. After a certain temperature, according to our data about 15 degrees celsuis, as temperature increases humidity decreases. Density altitude could occur at the points above the loess fit line where the temperature and humidity are both high.

Below is the same information except faceted by month. This plot shows us which months are most “at risk” for high temperatures and high humidity at LAX.

ggplot(cleandata, aes(x = temp, y = rh^3)) + geom\_point() + geom\_smooth(method = "loess") +   
 facet\_wrap(~month, nrow = 3) + xlab("Temperature (Celsius)") + ylab("Relative Humidity")

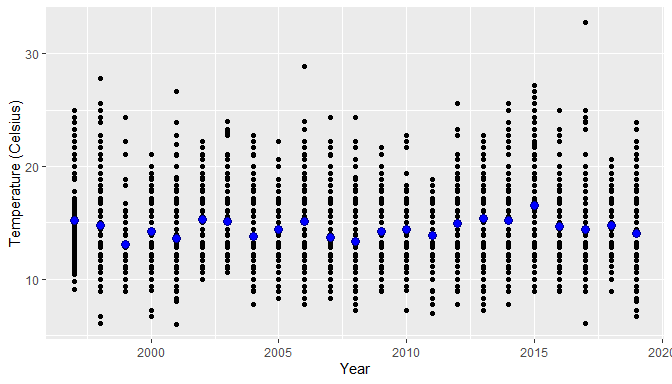


July, August, and September are the most at risk for density alitutde because of their high values along the x axis.

### Temperature

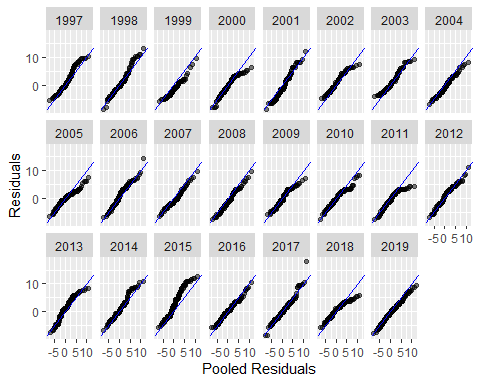
we will now look at temperature as a function of time to see if there has been any significant change. The following graph shows the temperature in celsius across different years. The blue point represents the average temperature for the year.

ggplot(cleandata, aes(x = year, y = temp)) + geom\_point() +  
 ylab("Temperature (Celsius)") + xlab("Year") +   
 stat\_summary(fun = "mean", geom = "point", cex = 3, pch = 21, col = "DarkBlue", bg="blue")



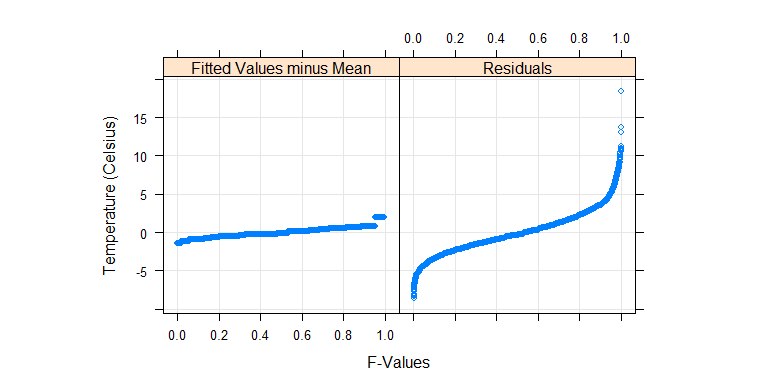
On first appearances there does not appear to be significant difference between the groups, but to be sure, we will look at a residual fit plot to determine if the variance is caused by random chance or if year could cause the variation. Before doing this, we must make sure the spreads between groups are relatively similar.

# get residuals, fvalue, and pooled residuals of temperature  
rftemp <- cleandata %>%   
 group\_by(year) %>%   
 mutate(residuals = temp - mean(temp)) %>%   
 dplyr::group\_by(year) %>%  
 dplyr::arrange(residuals) %>%   
 dplyr::mutate(f.val = (row\_number(residuals) - 0.5) / n() ) %>%  
 ungroup() %>%  
 mutate(Pooled.res = quantile(residuals, probs = f.val)) %>%  
 select(year, residuals, Pooled.res)  
  
# spreads essentially have the same spread so we can compare batches in residual-fit plot  
ggplot(rftemp, aes(y = residuals, x = Pooled.res)) + geom\_point(alpha = 0.5) +   
 geom\_abline(intercept = 0, slope = 1, color = "blue") +  
 facet\_wrap(~ year, nrow = 3) + xlab("Pooled Residuals") + ylab("Residuals")



The pooled residual plot shows that the batches essentially have the same spread except for 2000, 2011, 2015, and 2018. As the sample size of groups were small, this is okay, and we can conclude they all essentially had the same spreads. Because the spreads are similar across batches, we can now do a residual fit plot.

rfs(oneway(temp ~ year, data = cleandata, spread = 1),   
 aspect = 1,   
 ylab = "Temperature (Celsius)", xlab = "F-Values")



This residual - fit plot shows the temperature data normalized to the global mean against the residuals, with generated f-values. It reveals that the spread of the fitted teperatures across each year is insignificant compared to the spread of the residuals, so temperature differences could be explained by random chance. We will now do an Anova test to see if there are any years that are statistically different.

tempregression <- lm(temp ~ year, cleandata)  
anova(tempregression)

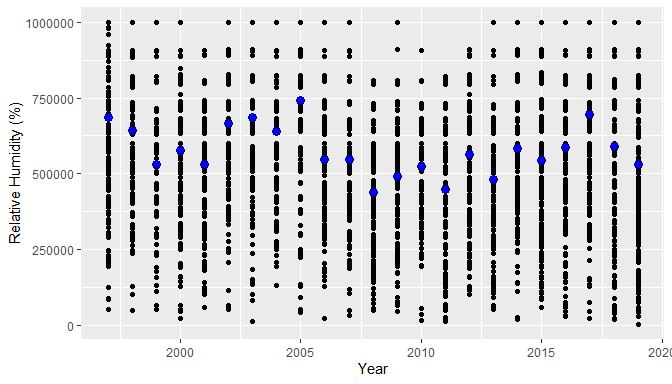
## Analysis of Variance Table  
##   
## Response: temp  
## Df Sum Sq Mean Sq F value Pr(>F)   
## year 1 37 36.881 4.1027 0.04286 \*  
## Residuals 5363 48210 8.989   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

With α = .05, a p-value of 0.1316 means that we can reject the null hypothesis and cannot conclude whether or not any of the groups’ mean are statistically significant.

### Humidity

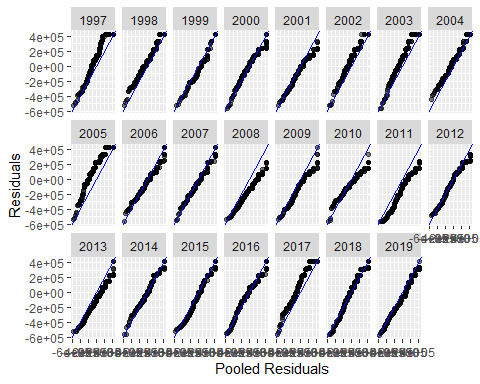
We will now analyze relative humidity in the same way. First we begin with a simple plot of relative humidity as a function of time.

ggplot(cleandata, aes(x = year, y = rh^3)) + geom\_point() +  
 ylab("Relative Humidity (%)") + xlab("Year") +   
 stat\_summary(fun = "mean", geom = "point", cex = 3, pch = 21, col = "DarkBlue", bg="blue")



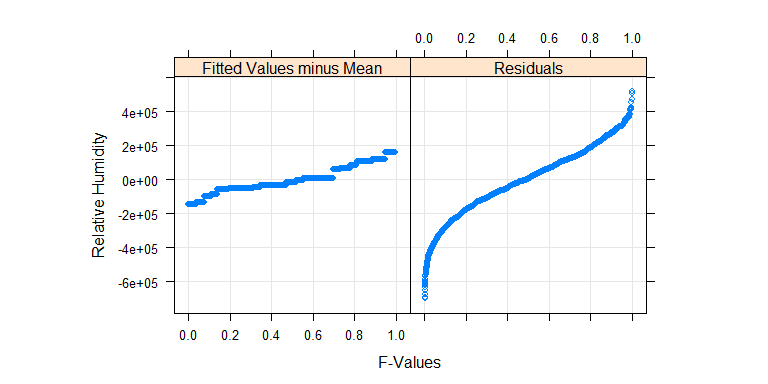
Unlike with temperature, relative humidity does appear to be changing throughout time; the mean temperature in blue is slightly decreasing throughout the years.

rfrh <- cleandata %>%   
 group\_by(year) %>%   
 mutate(residuals = rh^3 - mean(rh^3)) %>%   
 dplyr::group\_by(year) %>%  
 dplyr::arrange(residuals) %>%   
 dplyr::mutate(f.val = (row\_number(residuals) - 0.5) / n() ) %>%  
 ungroup() %>%  
 mutate(Pooled.res = quantile(residuals, probs = f.val)) %>%  
 select(year, residuals, Pooled.res)  
  
# spreads essentially have the same spread so we can compare batches in residual-fit plot  
ggplot(rfrh, aes(y = residuals, x = Pooled.res)) + geom\_point(alpha = 0.5) +   
 geom\_abline(intercept = 0, slope = 1, color = "blue") +  
 facet\_wrap(~ year, nrow = 3) + xlab("Pooled Residuals") + ylab("Residuals")



The spreads between matches are consistent enough, with the sample size of n = 235, to be able to do a residual fit plot.

rfs(oneway(rh^3 ~ year, data = cleandata, spread = 1),   
 aspect = 1,   
 ylab = "Relative Humidity", xlab = "F-Values")

 The residual fit plot shows that relative humidity was not significant in relation to the residuals; this means that the variation between years could be explained by random chance.

tempregression <- lm(rh^3 ~ year, cleandata)  
anova(tempregression)

## Analysis of Variance Table  
##   
## Response: rh^3  
## Df Sum Sq Mean Sq F value Pr(>F)   
## year 1 2.4586e+12 2.4586e+12 49.89 1.833e-12 \*\*\*  
## Residuals 5363 2.6429e+14 4.9281e+10   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

With α = .05, a p-value of 1.833 x 10-12 means that we cannot reject the null hypothesis, and we can conclude that mean relative humidity across the years do statistically differ from one another.

# Conclusion

Humidity and temperature are both highest in the months of July, August, and September. These are the months to be most aware of the issue of density altitude. While it does not appear that mean temperature has changed throughout the years. It does appear that mean humidity has changed significantly in the area surrounding Los Angeles Airport in a negative pattern.

# References

1. Dotson, Dianne J. “How Temperature & Humidity Are Related.” Sciencing, 23 April. 2018, <sciencing.com/temperature-ampamp-humidity-related-7245642.html>
2. “Three Hs of Aircraft Performance - Hot, High and Humid.” Desert Jet, 23 Aug. 2011, <www.desertjet.com/2011/08/23/aircraft-performance/>