Analyzing Heat and Humidity at LA International Airport

Makaylah Cowan

5/1/2020

Table of Contents

[Introduction 1](#_Toc39853019)

[Methods 2](#_Toc39853020)

[Data 2](#_Toc39853021)

[Software 2](#_Toc39853022)

[Assumptions 2](#_Toc39853023)

[Results 6](#_Toc39853024)

[Correlation 6](#_Toc39853025)

[Temperature 8](#_Toc39853026)

[Humidity 11](#_Toc39853027)

[Conclusion 13](#_Toc39853028)

[References 13](#_Toc39853029)



Aircraft Taking Off on Sunny Day Source: WallpaperSafari.

# Introduction

Relative humidity is the percentage of water vapor in the air. For reference, really saturated air that cannot hold any more water would have a relative humidity of 100%. If air’s temperature increases then it is able to 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. Keeping this in mind, 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 humidity, and analyze how the variables have changed individually at Los Angeles international airport across the years in order to see if there are more instances of high heat, humid days.

# Methods

### Data

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 1972 and 2020 were removed because they did not have complete, annual observations.

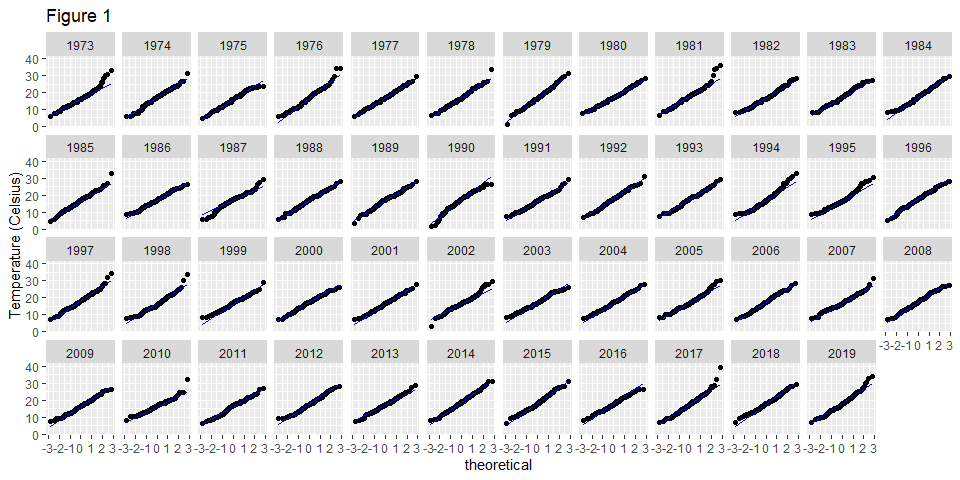
### Software

All analysis were done in R using dplyr, plyr, ggplot2, lubridate, and lattice packages.

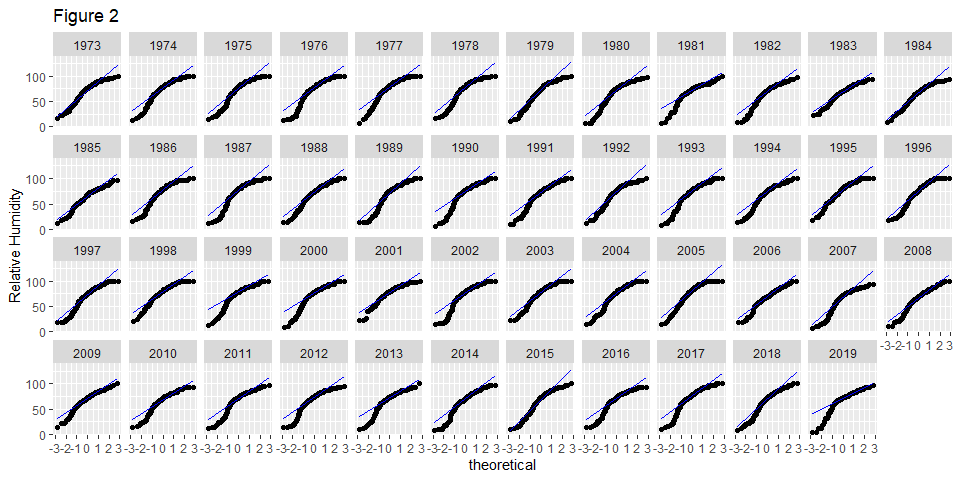
### Assumptions

# Setup  
library(dplyr)  
library(plyr)  
library(ggplot2)  
library(lubridate)  
library(lattice)  
  
Airport <- readRDS("../Data/LA\_airport.rds")  
  
cleandata <- Airport %>%   
 na.omit() %>%   
 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$year==1972),]  
cleandata <- cleandata[!(cleandata$year==2020),]  
  
cleandata2 <- cleandata  
cleandata <- sample\_n(cleandata,10000)

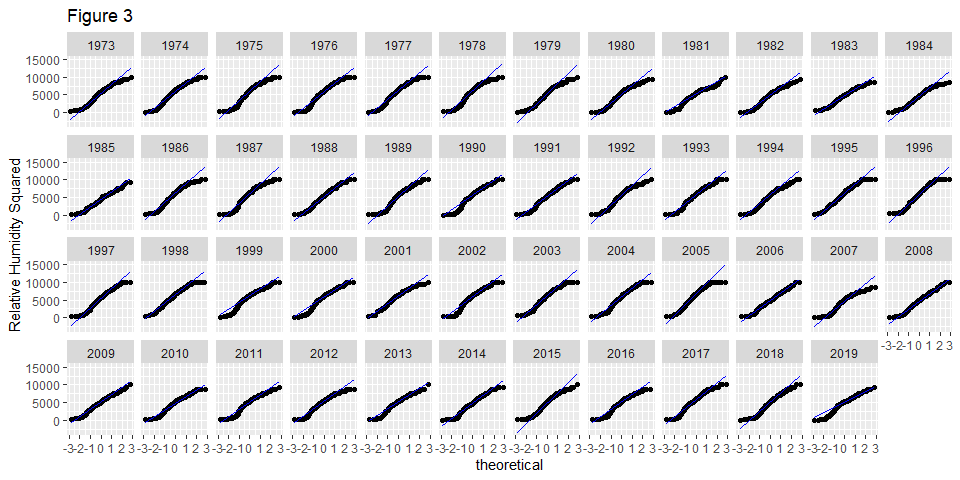
ggplot(cleandata, aes(sample = temp)) + geom\_qq(distribution = qnorm) +   
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "blue") + ylab("Temperature (Celsius)") +  
 facet\_wrap(~ year, nrow = 4) + ggtitle("Figure 1")



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 = 4) + ylab("Relative Humidity") +ggtitle("Figure 2")



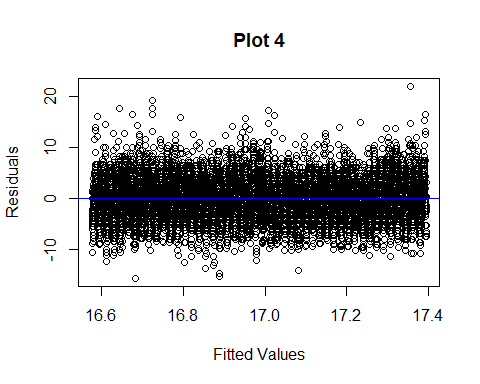
ggplot(cleandata, aes(sample = rh^2)) + geom\_qq(distribution = qnorm) +   
 geom\_qq\_line(line.p = c(0.25, 0.75), col = "blue") + ylab("Relative Humidity Squared") +  
 facet\_wrap(~ year, nrow = 4) + ggtitle("Figure 3")



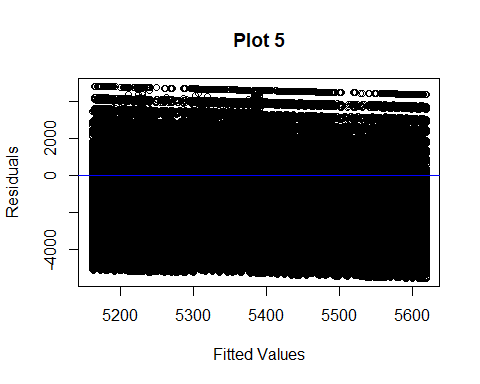
For the use of anova tests, the temperature and relative humidity variables must have sample independence, normality, and equality of variances. The sample independence requirement is met based on the experimental setup. Figure 1, a theoretical quantile-quantile plot, shows that temperature data appears to generally follow the distribution line, suggesting normality. The years 1978, 1981, 1989, 2005, 2010, and 2017 have some endpoints straying from the line, but this should not dramatically affect the results.

Plot 2 shows that the humidity data has extreme division from the line, and the shape of the endpoints suggests left skewed data; this meaning that the median value is greater than the mean value. To satisfy the normality requirement of later tests, the relative humdity data will be expressed by being squared as in Plot 3.

tempmod <- lm(temp~time,cleandata)  
plot(tempmod$fitted.values,tempmod$residuals, main = "Plot 4", xlab= "Fitted Values", ylab = "Residuals")  
abline(h=0, col="blue")



rhmod <- lm(rh^2~time,cleandata2)  
plot(rhmod$fitted.values,rhmod$residuals, main = "Plot 5", xlab= "Fitted Values", ylab = "Residuals")  
abline(h=0, col="blue")

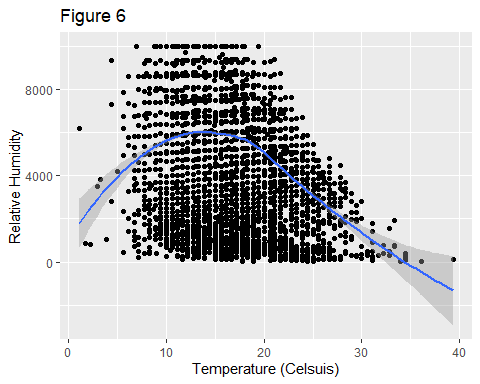


Plots 4 and 5 show that there is no apparent trend or pattern in the fitted values vs. residuals for neither the temperature nor the relative humdity data. This suggestst that there is homogeneity in variance for each variable.

# Results

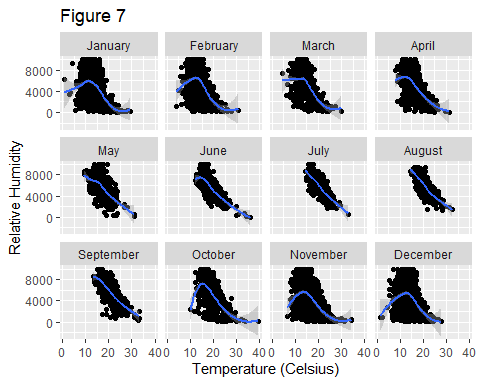
### Correlation

ggplot(cleandata, aes(x = temp, y = rh^2)) + geom\_point() + geom\_smooth(method = "loess") + xlab("Temperature (Celsius)") +  
 ylab("Relative Humidity") + ggtitle("Figure 6")



As expected, the relationship between temperature and re-expressed humidity values is direct and negative. Beyond a temperature range, around 15 degrees celsius, 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.

ggplot(cleandata, aes(x = temp, y = rh^2)) + geom\_point() + geom\_smooth(method = "loess") +   
 facet\_wrap(~month, nrow = 3) + xlab("Temperature (Celsius)") + ylab("Relative Humidity") + ggtitle("Figure 7")



The same data except faceted by month shows us which months are most “at risk” for high temperatures and high humidity at LAX. Relative to their number of samples, July, August, and September are the most at risk months because of their temperature values and the number of points above the loess fit line.

### Temperature

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") + ggtitle("Figure 8")

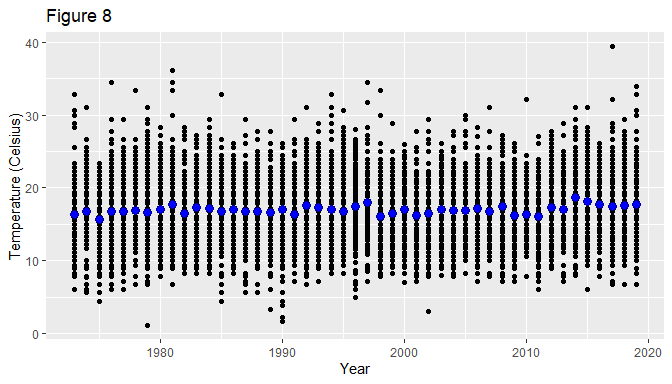
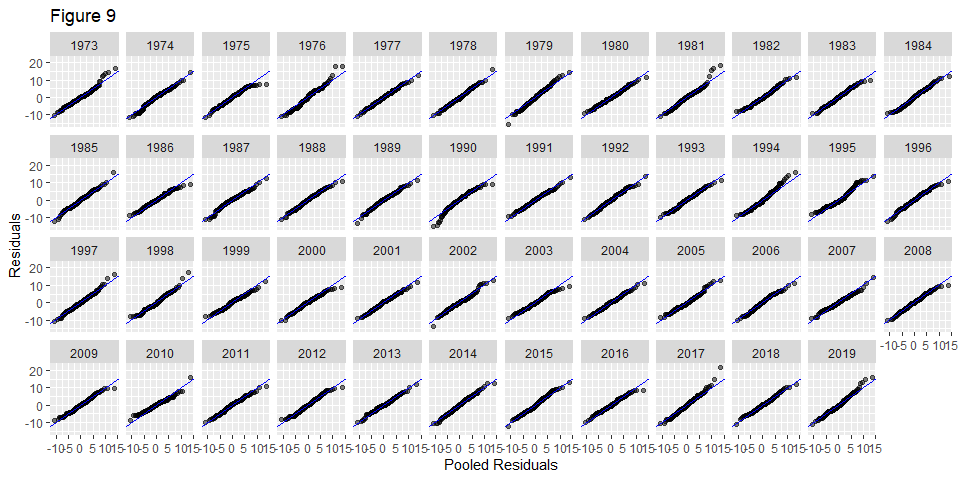


Figure 8 shows temperature as a function of time in years. The blue point represents the average temperature for the each year. On first appearances there does not appear to be significant difference between the groups. There does appear to be a slight upward trend; this will be checked by looking at a residual fit plot to determine if variance is caused by random chance.

# 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 = 4) + xlab("Pooled Residuals") + ylab("Residuals") + ggtitle("Figure 9")



The pooled residual plot shows that the batches essentially have the same spread, so it is appropriate to do a residual fit plot.

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

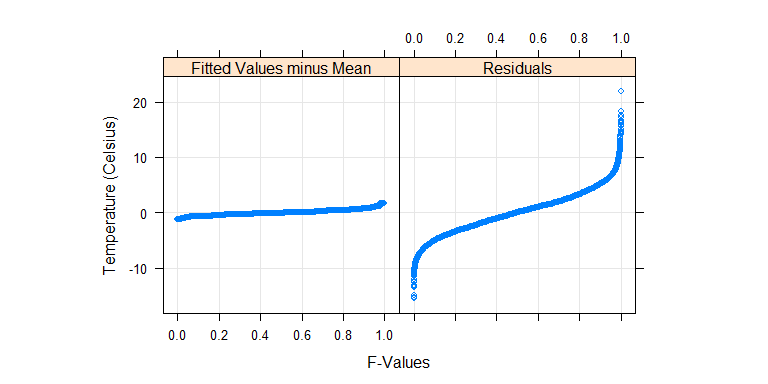


Figure 10 shows the temperature data normalized to the global mean against the residuals, with generated f-values. It reveals that the spread of the fitted teperatures across each year is insignificant compared to the spread of the residuals, so temperature differences could be explained by random chance. An Anova test will reveal if any years that are statistically different.

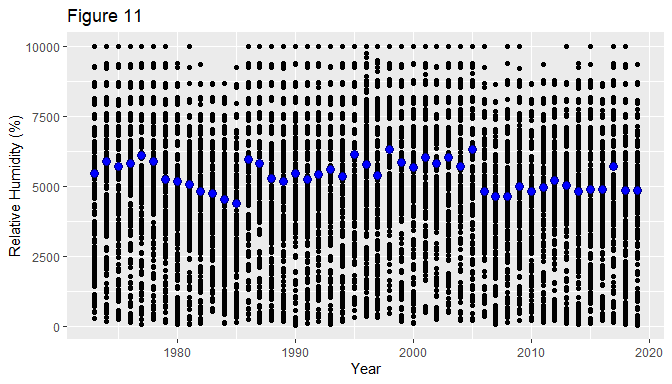
tempregression <- lm(temp ~ year, cleandata2)  
anova(tempregression)

## Analysis of Variance Table  
##   
## Response: temp  
## Df Sum Sq Mean Sq F value Pr(>F)   
## year 1 17389 17389.4 1058.3 < 2.2e-16 \*\*\*  
## Residuals 409256 6724591 16.4   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

With α = .05, a p-value of 2.2 x 10-16 means that we can reject the null hypothesis and can conclude that some of the temperature means do differ statistically across the years.

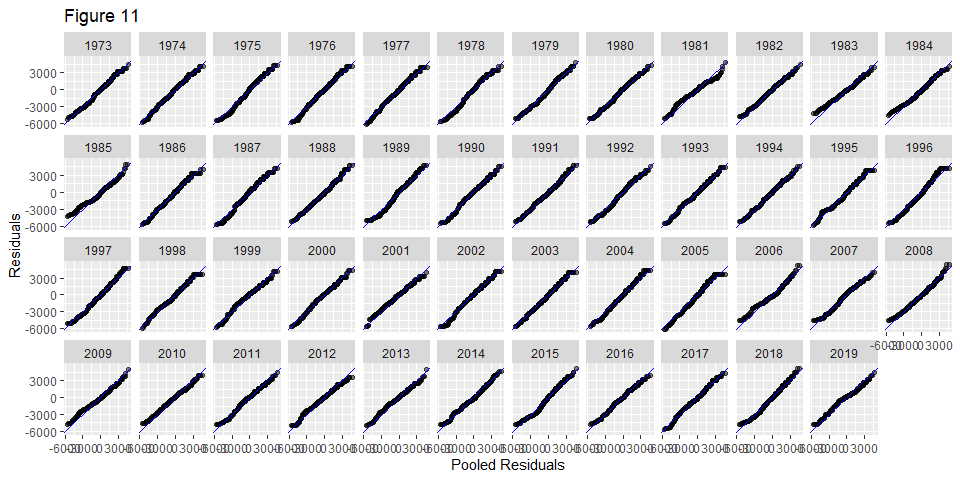
### Humidity

ggplot(cleandata, aes(x = year, y = rh^2)) + geom\_point() +  
 ylab("Relative Humidity (%)") + xlab("Year") +   
 stat\_summary(fun = "mean", geom = "point", cex = 3, pch = 21, col = "DarkBlue", bg="blue") + ggtitle("Figure 11")



Unlike with temperature, relative humidity does not appear to have a consistent trend; the mean temperature in blue varies wildly but appears to be decreasing overall.

rfrh <- cleandata %>%   
 group\_by(year) %>%   
 mutate(residuals = rh^2 - mean(rh^2)) %>%   
 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 = 4) + xlab("Pooled Residuals") + ylab("Residuals") + ggtitle("Figure 11")



rfs(oneway(rh^2 ~ year, data = cleandata, spread = 1),   
 aspect = 1,   
 ylab = "Relative Humidity", xlab = "F-Values")

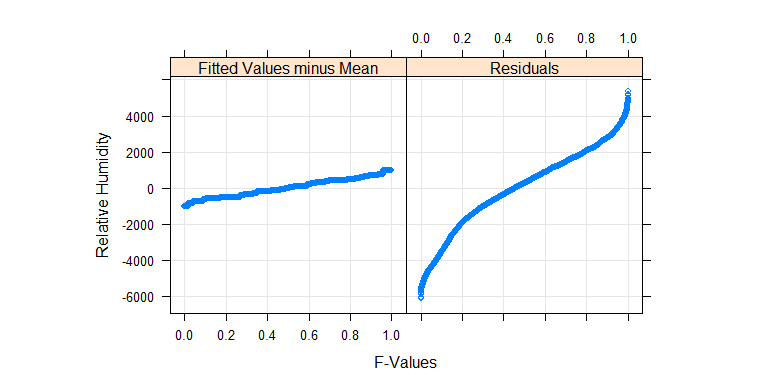


Figure 11 shows that the spreads between matches are consistent enough to be able to do a residual fit plot. Subsequently, figure 12 shows that relative humidity was not significant in relation to the residuals; this means that the variation between years could be explained by random chance.

rhregression <- lm(rh^2 ~ year, cleandata2)  
anova(rhregression)

## Analysis of Variance Table  
##   
## Response: rh^2  
## Df Sum Sq Mean Sq F value Pr(>F)   
## year 1 6.9852e+09 6985186722 1267 < 2.2e-16 \*\*\*  
## Residuals 409256 2.2562e+12 5512955   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

With α = .05, a p-value of 2.2 x 10-16 means that we can 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, so these are the months to be most aware of the issue of density altitude. Across time, the trend for temperature visually appears to increase while there is not an immediately recognizable trend for humidity. However, both variables have p-values of 2.2 x 10-16 which means that they have both changed signficantly across the years. Increased awareness is needed for this issue, especially in the identified months, as these variables continue to change.

# References

1. Dotson, Dianne J. “How Temperature & Humidity Are Related.” Sciencing, 23 April. 2018, [sciencing.com/temperature-ampamp-humidity-related-7245642.html](file:///C:\tmp\es218_project\Presentation\sciencing.com\temperature-ampamp-humidity-related-7245642.html)
2. Gimond, Manuel. Exploratory Data Analysis in R. [<http://mgimond.github.io/ES218/index.html>] (<http://mgimond.github.io/ES218/index.html>)
3. “Three Hs of Aircraft Performance - Hot, High and Humid.” Desert Jet, 23 Aug. 2011, [www.desertjet.com/2011/08/23/aircraft-performance/](file:///C:\tmp\es218_project\Presentation\www.desertjet.com\2011\08\23\aircraft-performance\)
4. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.