An Analysis:

The Effect of Humid Environments on Human Performance

(measured using non-invasive running metrics from widely available wrist based consumer devices)

University of Louisiana at Lafayette

INFX 595

Dr. Mehmet Tozal

Damian O'Boyle
(C00481724)

Submitted: March 2022

Table of Contents

Introduction

The different regions of the world offer their locales contrasting weather climates throughout a typical calendar year. Areas closer to the equator tend to record higher temperatures more often than those at more northerly latitudes. For islands/regions surrounded by large bodies of water that have a greater availability to moisture, these higher temperatures add to the air's ability to hold increased levels of that available moisture often leading to humid or 'muggy' feeling conditions. The higher the air temperature the greater its moisture holding ability. The aim of this analysis is to understand whether this increased level of relative humidity has a measurable effect on human performance compared with colder or less humid regions/environments.

This report outlines the steps taken in an exploratory analysis of the effect of humidity on high performance athletics using metrics from wrist based consumer fitness products (typically from Garmin®). As such, further, more in-depth follow-up opportunities for research will be available and necessary in order to fully understand the hypotheses posed. The entire project hinges on the suspected hypothesis that humid environments take a greater physical toll on the body while engaging in high performance athletics, that is to say that they have a negative performance effect, when compared with that of less humid or 'nominal' environmental conditions. This hypothesis is postured based on real world experience, feedback and reports from athletes who have trained and competed across these differing climates.

Dataset

Description

This is a proprietary dataset compiled from scratch specifically for the purpose of this data analysis project. The specific data points relate to easily obtained (non-invasive) performance based running metrics which were recorded using wrist based consumer fitness products. These metrics include heart rate, pace and elevation climbed during one mile segments of running, and are coupled with popular weather based metrics including temperature, humidity, wind speed and pressure etc. The data points are suspected to prove useful in understanding the effect of humid environments on high performance activity.

There are a total of eight data subjects that have contributed to this dataset, each offering seventeen different and unique variable attributes (columns) in this dataset, with insight into between 669-1421 individual unique mile segments (rows).

Data Origin

The data used to compile the dataset were manually transcribed using Microsoft Excel®, from publicly available sources present online at Strava.com (running based metrics) and weatherunderground.com (weather based metrics). The data available from these sources was thoughtfully selected for use in this analysis. The collection process began, starting in September of 2021. However, due to the unforeseen scale and effort involved in manual transcription of a large majority of this data, the process continued until early March of 2022.

Data Cleaning

The use of Microsoft Excel® to compile the initial raw data allowed for quick and easy manipulation into the neat, efficient dataset eventually used and presented. Manual transcription allowed for preemptive cleaning of unnecessary metadata as well as perceived outliers. (The average heart rate value recorded in the first mile segment of each activity was removed as the data subjects would begin their workouts from a resting or near resting heart rate which would typically take anywhere between half and a full mile to reach a steady rate).

The result, eight individual datasets consisting of thirteen different variable fields. Through calculation other variables were added to the weather based data, Heat Index (HI) being a continuous example and DewFeel and RealFeel being categorical examples. The precipitation field was ultimately removed as this field almost exclusively provided a zero value. (This was suspected to have been an anomaly caused by the nature of the historic weather data that was available, coming exclusively from airport stations, where wind readings are typically the primary concern).

The individual datasets were then combined into one, adding name and gender identifiers to aid subsetting during analysis. The final configuration resulted in a single dataset consisting of seventeen variables, representing a total of 7,922 observations. Precautions were taken to preserve and protect the identities of those who contributed their data.

Table of Variables

Name	Mode	Description
Name	factor	An identifier for each individual data subject
Gender	factor	The data subjects assigned gender, denoted as M (Male) or F (Female)
Date	date	The local date on which the activity was conducted (dd/mm/yyyy) format
Time	character	The local time at which the activity was conducted (hh:mm) format
Location	character	The location in which the activity was conducted
Pace	integer	The average pace measured in seconds for the specific mile split recorded
HR	integer	The average heart rate recorded for the specific mile split recorded
Elevation	integer	The elevation climbed in feet during the specific mile split recorded
Temp	integer	The temperature measured in fahrenheit recorded during the activity
Humidity	integer	The relative humidity measured in percentage recorded during the activity
DP	integer	The dew point measured in fahrenheit calculated during the activity
HI	integer	The heat index measured in fahrenheit calculated during the activity
DewFeel	factor	An identifier used to denote how the dew point makes the air feel
RealFeel	factor	An identifier used to denote the risk level of the heat index on the body
Wind	integer	The average wind speed in miles per hour recorded during the activity
Gust	integer	The maximum wind speed in miles per hour recorded during the activity
Pressure	integer	The average air pressure measured during the activity in millibars

Data Loading

The cleaned dataset was loaded into R and structured using the following commands;

```
> dataset <- read.csv("C:\\...\\dataset.csv")
```

The structure of the dataset is shown;

```
> str(dataset)
```

```
'data.frame':    7922 obs. of  17 variables:
 $ Name      : chr  "Matt" "Matt" "Matt" "Matt" ...
 $ Gender    : chr  "M" "M" "M" "M" ...
 $ Date      : chr  "07/08/2018" "07/08/2018" "07/08/2018" "07/08/2018" ...
 $ Time      : chr  "09:05" "09:05" "10:10" "10:10" ...
 $ Location  : chr  "Waverly, England" "Waverly, England" "Waverly, England" ...
 $ Pace      : int  "399" "406" "456" "430" ...
 $ HR        : int  141 145 146 140 134 139 145 143 145 148 ...
 $ Elevation: int  -41 -30 56 -66 -100 80 -3 16 -52 49 ...
 $ Temp      : int  74 74 79 79 79 79 79 63 63 63 ...
 $ Humidity  : int  56 56 47 47 51 51 51 64 64 64 ...
 $ DP        : int  57 57 57 57 60 60 60 51 51 51 ...
 $ HI        : int  74 74 79 79 79 79 79 62 62 62 ...
 $ DewFeel   : chr  "Comfortable" "Comfortable" "Comfortable" ...
 $ RealFeel  : chr  "Okay" "Okay" "Okay" "Okay" ...
 $ Wind      : int  2 2 5 5 8 8 8 9 9 9 ...
 $ Gust      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Pressure  : int  1001 1001 1001 1001 999 999 999 1004 1004 ...
```

The highlighted attributes were converted to factor data types. The following commands were used to achieve this;

```
> dataset$Name <- as.factor(dataset$Name)
> dataset$Gender <- as.factor(dataset$Gender)
> dataset$DewFeel <- as.factor(dataset$DewFeel)
> dataset$RealFeel <- as.factor(dataset$RealFeel)
```

The head command was used as follows to display the first six columns of the dataset.

```
> head(dataset)
```

	Name	Gender	Date	Time	Location	Pace	HR	Elevation	Temp	Humidity	DP	HI	DewFeel	RealFeel	Wind	Gust	Pressure
1	Matt	M	07/08/2018	09:00	Waverly, England	399	141	-41	74	56	57	74	Comfortable	Okay	2	0	1001
2	Matt	M	07/08/2018	09:00	Waverly, England	406	145	-30	74	56	57	74	Comfortable	Okay	2	0	1001
3	Matt	M	07/08/2018	10:00	Waverly, England	456	146	56	79	47	57	79	Comfortable	Okay	5	0	1001
4	Matt	M	07/08/2018	10:00	Waverly, England	430	140	-66	79	47	57	79	Comfortable	Okay	5	0	1001
5	Matt	M	07/08/2018	16:00	Waverly, England	436	134	-100	79	51	60	79	Comfortable	Okay	8	0	999
6	Matt	M	07/08/2018	16:00	Waverly, England	424	139	80	79	51	60	79	Comfortable	Okay	8	0	999

The data frame was then divided as necessary into different lists based on Gender and Name. These splits allow for greater depth of analysis into the available data.

```
> gender <- split(dataset, dataset$Gender)
> subject <- split(dataset, dataset$Name)
```

The above splits return similar outputs for the above functions and so are not shown.

Expectations

Firstly, it is expected that correlations will be easily found between the weather based metrics present in the dataset. These metrics are closely related to, and often derived from one another so a lack of correlation would actually be more surprising.

Through analysis of the available data it is suspected that mile segments (observations) recorded in humid conditions will present a variety of effects on the running based metrics Heart Rate (HR) and Pace. It is expected that higher HR values will be recorded when segments are completed in humid conditions compared to more nominal or 'dry' conditions. Indicating that subjects are working harder. A similar outcome is expected for the Pace variable, slower paces are expected in humid conditions or more specifically slower paces in respect to HR values are expected when compared with segments recorded in nominal humidity conditions. These expectations will likely be best tested through regression analysis. It is expected, through this analysis, that the typical effect on HR or Pace, caused by, for example, a change in Dew Point (DP) metric can be calculated.

HR and Pace data are expected to be normally distributed, for obvious reasons. Athletes are more likely to typically run within a comfort zone during everyday workouts, exerting more effort when conducting interval or tempo sessions etc. and taking it easier during warmups or cooldowns etc. both of which occur far less frequently than the daily run.

Dew Point is considered the most correct methodology to measure the effect or feel of humidity. The metric relating to relative humidity (Humidity in the dataset) is a percentage that is calculated based on the amount of moisture in the air over the total moisture holding capacity of the air which increases with a rise in the air temperature. For this reason relative humidity is often unreliable as a metric and difficult for human observers to interpret. Whereas dew point on the other hand is an independent value not relying on any other metric for its interpretation.

Therefore, DP and Heat Index (HI) are expected to feature heavily in predicting performance effects on the HR and Pace, these values will also incur the effect of Temp and Humidity, as these variables are all interrelated in their methodology of calculation.

Limitations

The weather archive (WeatherUnderground) used to compile the dataset, appeared during the data collection phase to have issues differentiating between hours during daylight savings time. This was noticed when the time was reverted back to Universal Time in October 2021. The measurements for some of the already compiled datasets at that time appeared to have shifted an hour, this may also have been a manual transcription error or a difference of interpretation in which hour measurement to use. These datasets were fixed to align with the Universal Time at their respective locations, however for this reason some of the data for segments recorded in spring/summer months where Daylight Saving Time is observed may be for an incorrect earlier hour. The dataset was completed before the time changed to Daylight Savings Time again in March 2022. So the issue of the recorded weather data shifting again should have been negated.

The archived weather readings were all collected from airport weather stations as these are required to make hourly or more frequent records of weather conditions and keep a historic log to aid with aircraft control. These areas are open and therefore typically more prone to higher wind measurement readings. They can also be a bit further from population centres where activities were actually conducted. This may be a contributing factor to data variability.

Due to the energy sapping heat and other difficult conditions, athletes tend to run at earlier or later hours when the temperature is lower. For this reason the full extent of the effect imposed by the heat and humidity may be more difficult to truly understand if it were to be compared with the most extreme conditions available within each day closer to solar noon.

Finally, the dataset itself represents a relatively small sample size of available data. For this reason the analysis conducted may not be representative of the greater population at large. The small size may also unproportionally skew the results returned here, due to the high likelihood of variability in the data which was seen during collection. Skewing is expected to be more likely encountered with the subsets representing Gender or individual data subjects.

Analysis

Correlations

```
> cor(dataset[c(6:12, 15:17)]
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.00000000	-0.107500938	0.0523250170	0.0200951860	0.123513069	0.088010139	0.0274589433	-0.0296696236	-0.029021501	0.05541798
HR	-0.10750094	1.000000000	0.0080526776	-0.1296738162	0.111794170	-0.054050230	-0.1241814777	-0.0242996947	-0.023353011	0.02397881
Elevation	0.05232502	0.008052678	1.0000000000	-0.0003936188	-0.009133338	-0.008565611	-0.0009634305	-0.0003045178	0.001286247	-0.01050534
Temp	0.02009519	-0.129673816	-0.0003936188	1.0000000000	-0.182228061	0.776895372	0.9979469355	-0.0804805327	-0.032341784	-0.08458688
Humidity	0.12351307	0.111794170	-0.0091333378	-0.1822280612	1.000000000	0.460535351	-0.1214626979	-0.1846738842	-0.118259969	0.34584298
DP	0.08801014	-0.054050230	-0.0085656115	0.7768953720	0.460535351	1.000000000	0.8126658593	-0.1589420721	-0.086961575	0.19927554
HI	0.02745894	-0.124181478	-0.0009634305	0.9979469355	-0.121462698	0.812665859	1.0000000000	-0.0935350430	-0.040079318	-0.06519627
Wind	-0.02966962	-0.024299695	-0.0003045178	-0.0804805327	-0.184673884	-0.158942072	-0.0935350430	1.0000000000	0.505676823	0.08024366
Gust	-0.02902150	-0.023353011	0.0012862471	-0.0323417836	-0.118259969	-0.086961575	-0.0400793185	0.5056768233	1.000000000	-0.01032800
Pressure	0.05541798	0.023978812	-0.0105053383	-0.0845868802	0.345842982	0.199275543	-0.0651962701	0.0802436551	-0.010328000	1.00000000

The values with a level of significance higher than 0.5 have been highlighted for easy viewing (Green: > ±0.9, yellow: > ±0.8, orange: > ±0.7, red: > ±0.6, blue: > ±0.5).

There were only three significant observations above the threshold level of ±0.7 that can be gleaned from this correlation test, one at each threshold level.

Temp/HI	0.9979469355
HI/DP	0.8126658593
DP/Temp	0.7768953720

These three correlations are all weather based metrics which are all closely related and/or derived from each other. Therefore their correlation was expected and is ultimately of little use to this analysis, other than offering support to the belief that the data is accurate in relation to itself.

On the other hand, the lack of correlated values in the dataset specifically related to the performance metrics, Pace and HR with respect to the weather based variables, is surprising and may contribute to this analysis proving difficult to conduct. It is worth remembering that these are the correlation figures for the entire dataset which is made up of eight different data subjects, four of each gender. The gender specific correlations and the data subjects are hoped to prove more lucrative, as even the HR and Pace variables are not strongly correlated here.

```
> cor(gender$F[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.000000000	-0.01645177	0.035202415	0.14496148	0.17122674	0.26311121	0.15746906	-0.003748913	0.02105844	0.214925879
HR	-0.016451765	1.00000000	-0.025384633	0.04117318	0.11972440	0.11525625	0.04866172	0.052113668	0.01583695	0.091375214
Elevation	0.035202415	-0.02538463	1.000000000	0.01270601	-0.01308119	0.00261795	0.01197146	-0.012257455	-0.01395805	-0.005705767
Temp	0.144961477	0.04117318	0.012706013	1.00000000	-0.22906746	0.74217810	0.99776883	-0.149620323	-0.08765618	-0.079805920
Humidity	0.171226740	0.11972440	-0.013081187	-0.22906746	1.00000000	0.46244737	-0.16610261	-0.104948374	-0.05284233	0.454797992
DP	0.263111209	0.11525625	0.002617950	0.74217810	0.46244737	1.00000000	0.78195477	-0.171096560	-0.08854227	0.283736341
HI	0.157469060	0.04866172	0.011971458	0.99776883	-0.16610261	0.78195477	1.00000000	-0.160140249	-0.09134863	-0.052708024
Wind	-0.003748913	0.05211367	-0.012257455	-0.14962032	-0.10494837	-0.17109656	-0.16014025	1.000000000	0.43765570	0.083144426
Gust	0.021058442	0.01583695	-0.013958049	-0.08765618	-0.05284233	-0.08854227	-0.09134863	0.437655696	1.00000000	0.082508538
Pressure	0.214925879	0.09137521	-0.005705767	-0.07980592	0.45479799	0.28373634	-0.05270802	0.083144426	0.08250854	1.000000000

```
> cor(gender$M[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.00000000	-0.506532049	0.064613948	0.140932031	-0.04901609	0.09233269	0.137364256	0.041621764	0.020617178	-0.016851406
HR	-0.50653205	1.000000000	0.018424834	-0.150956595	0.03658275	-0.10888806	-0.149595343	-0.022960720	-0.001306715	-0.000533967
Elevation	0.06461395	0.018424834	1.000000000	-0.003476761	-0.01052503	-0.01253140	-0.004121065	0.007593914	0.009330232	-0.012401374
Temp	0.14093203	-0.150956595	-0.003476761	1.000000000	-0.09474600	0.79970669	0.998012035	-0.070817789	-0.033215674	-0.099318363
Humidity	-0.04901609	0.036582753	-0.010525031	-0.094746003	1.00000000	0.50813215	-0.034566055	-0.225675037	-0.142203906	0.284557079
DP	0.09233269	-0.108888056	-0.012531397	0.799706685	0.50813215	1.00000000	0.833229871	-0.172458063	-0.105143908	0.143405055
HI	0.13736426	-0.149595343	-0.004121065	0.998012035	-0.03456606	0.83322987	1.000000000	-0.084673742	-0.042366689	-0.083594933
Wind	0.04162176	-0.022960720	0.007593914	-0.070817789	-0.22567504	-0.17245806	-0.084673742	1.000000000	0.537001165	0.075450686
Gust	0.02061718	-0.001306715	0.009330232	-0.033215674	-0.14220391	-0.10514391	-0.042366689	0.537001165	1.000000000	-0.059724834
Pressure	-0.01685141	-0.000533967	-0.012401374	-0.099318363	0.28455708	0.14340505	-0.083594933	0.075450686	-0.059724834	1.000000000

The male gender specific correlation table definitely offers a better insight into the data relationships that exist in this dataset, with increased rates of correlation between HR and Pace. However these values are still not near significant values, nor do they suggest any correlation to the desired weather based variables. The female specific table offers less useful insight than the correlations for the full dataset.

```
> cor(subject$A[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.00000000	-0.306440995	0.029788136	0.282639591	-0.041708998	0.22032050	0.281940163	-0.100763246	-0.05789543	-0.08365703
HR	-0.30644099	1.000000000	0.009231818	0.179459726	-0.039639170	0.14176018	0.173940610	0.009997739	0.05689252	-0.04505361
Elevation	0.02978814	0.009231818	1.000000000	-0.005896713	0.082669242	0.04155795	-0.001018506	-0.015054401	-0.01703089	0.06939646
Temp	0.28263959	0.179459726	-0.005896713	1.000000000	-0.066392694	0.84212460	0.997781728	-0.264488245	-0.09328860	-0.10134511
Humidity	-0.04170900	-0.039639170	0.082669242	-0.066392694	1.000000000	0.47578704	-0.002790171	-0.419196279	-0.14586617	0.01222782
DP	0.22032050	0.141760179	0.041557950	0.842124599	0.475787040	1.00000000	0.874010781	-0.444413611	-0.14712557	-0.07472379
HI	0.28194016	0.173940610	-0.001018506	0.997781728	-0.002790171	0.87401078	1.000000000	-0.294213705	-0.10002495	-0.09953713
Wind	-0.10076325	0.009997739	-0.015054401	-0.264488245	-0.419196279	-0.44441361	-0.294213705	1.000000000	0.43414114	-0.30890670
Gust	-0.05789543	0.056892522	-0.017030889	-0.093288597	-0.145866172	-0.14712557	-0.100024949	0.434141135	1.00000000	-0.06871432
Pressure	-0.08365703	-0.045053610	0.069396460	-0.101345111	0.012227820	-0.07472379	-0.099537126	-0.308906701	-0.06871432	1.00000000

```
> cor(subject$B[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.000000000	-0.404927497	0.087280152	0.345975111	-0.001753615	0.32853180	0.345156860	-0.07064102	0.057865298	0.194514366
HR	-0.404927497	1.000000000	-0.038736442	-0.040715122	0.067677043	-0.01237581	-0.036164267	-0.01808283	-0.009531145	0.058938823
Elevation	0.087280152	-0.038736442	1.000000000	0.021399853	-0.007628484	0.01660378	0.021009646	-0.01620004	-0.002780607	-0.012945836
Temp	0.345975111	-0.040715122	0.021399853	1.000000000	-0.051387170	0.92476735	0.998752515	-0.12939502	-0.007274546	0.342848620
Humidity	-0.001753615	0.067677043	-0.007628484	-0.051387170	1.000000000	0.32865519	-0.005294674	-0.46315669	-0.244534339	-0.239059464
DP	0.328531796	-0.012375812	0.016603780	0.924767351	0.328655191	1.00000000	0.940878325	-0.30087137	-0.100185996	0.239241112
HI	0.345156860	-0.036164267	0.021009646	0.998752515	-0.005294674	0.94087832	1.000000000	-0.15066727	-0.017231019	0.329472581
Wind	-0.070641015	-0.018082829	-0.016200040	-0.129395020	-0.463156690	-0.30087137	-0.150667274	1.00000000	0.393258409	-0.159311116
Gust	0.057865298	-0.009531145	-0.002780607	-0.007274546	-0.244534339	-0.10018600	-0.017231019	0.39325841	1.000000000	0.006867355
Pressure	0.194514366	0.058938823	-0.012945836	0.342848620	-0.239059464	0.23924111	0.329472581	-0.15931112	0.006867355	1.000000000

```
> cor(subject$C[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.00000000	-0.56032612	0.11877519	0.14554303	0.02088502	0.152251802	0.15024498	0.0343626699	0.059002593	-0.02426643
HR	-0.56032612	1.00000000	-0.03102719	0.07077196	-0.06461070	-0.012628685	0.06673017	-0.0147833035	-0.094554724	-0.08004550
Elevation	0.11877519	-0.03102719	1.00000000	0.02689645	-0.03157819	0.005324597	0.02483386	0.0008532401	0.002082585	0.03271713
Temp	0.14554303	0.07077196	0.02689645	1.00000000	-0.32241714	0.609039183	0.99720839	-0.0593407447	-0.114049890	-0.19380034
Humidity	0.02088502	-0.06461070	-0.03157819	-0.32241714	1.00000000	0.530381553	-0.25289738	0.0480180293	0.091285266	0.60836763
DP	0.15225180	-0.01262869	0.00532459	0.60903918	0.53038155	1.00000000	0.66283492	0.0082861592	-0.013033928	0.34940578
HI	0.15024498	0.06673017	0.02483386	0.99720839	-0.25289738	0.66283491	1.00000000	-0.0588345420	-0.109519382	-0.15274892
Wind	0.03436267	-0.01478330	0.00085324	-0.05934074	0.04801803	0.008286159	-0.05883454	1.0000000000	0.398146067	0.24899015
Gust	0.05900259	-0.09455472	0.00208258	-0.11404989	0.09128527	-0.013033928	-0.10951938	0.3981460675	1.000000000	0.13083421
Pressure	-0.02426643	-0.08004550	0.03271713	-0.19380034	0.60836763	0.349405776	-0.15274892	0.2489901452	0.130834213	1.00000000

```
> cor(subject$D[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.000000000	-0.31693966	0.021300659	0.23162155	-0.04401286	0.20585586	0.23168340	-0.230720467	-0.13095570	-0.004610111
HR	-0.316939656	1.00000000	0.025013886	0.14813264	0.10751329	0.19127981	0.15256962	0.120995744	-0.01748048	-0.233296886
Elevation	0.021300659	0.02501389	1.000000000	-0.02476146	0.01699601	-0.01858117	-0.02394595	0.007888779	-0.04002437	-0.093432469
Temp	0.231621553	0.14813264	-0.024761459	1.00000000	-0.18109239	0.91133514	0.99880911	-0.200797648	-0.11155650	0.114870489
Humidity	-0.044012860	0.10751329	0.016996010	-0.18109239	1.00000000	0.23483096	-0.13624189	-0.212522189	-0.28271946	-0.237058540
DP	0.205855864	0.19127981	-0.018581167	0.91133514	0.23483096	1.00000000	0.92861599	-0.283404080	-0.21369328	0.034070369
HI	0.231683398	0.15256962	-0.023945951	0.99880911	-0.13624189	0.92861599	1.00000000	-0.212298057	-0.12394905	0.102931712
Wind	-0.230720467	0.12099574	0.007888779	-0.20079765	-0.21252219	-0.28340408	-0.21229806	1.000000000	0.50823618	-0.276723117
Gust	-0.130955702	-0.01748048	-0.040024369	-0.11155650	-0.28271946	-0.21369328	-0.12394905	0.508236185	1.00000000	0.036765605
Pressure	-0.004610111	-0.23329689	-0.093432469	0.11487049	-0.23705854	0.03407037	0.10293171	-0.276723117	0.03676561	1.000000000

```
> cor(subject$E[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.000000000	-0.47286285	0.0972548863	0.073075338	-0.104363706	-0.007709416	0.0676214430	-0.004005673	0.075943828	-0.24744297
HR	-0.472862850	1.00000000	0.0184613546	-0.145651339	-0.031237413	-0.146016492	-0.1478324428	0.051908834	0.014828526	0.04137212
Elevation	0.097254886	0.01846135	1.0000000000	0.001059286	-0.009399605	-0.006763403	0.0004097364	-0.002607954	0.007255135	-0.01025430
Temp	0.073075338	-0.14565134	0.0010592862	1.000000000	-0.086602151	0.813183069	0.9982889857	-0.202012885	-0.105404883	-0.20531262
Humidity	-0.104363706	-0.03123741	-0.0093996046	-0.086602151	1.000000000	0.493515699	-0.0309170886	-0.244085773	-0.156574670	0.35598331
DP	-0.007709416	-0.14601649	-0.0067634030	0.813183069	0.493515699	1.000000000	0.8431035546	-0.283195935	-0.166324645	0.10230224
HI	0.067621443	-0.14783244	0.0004097364	0.998288986	-0.030917089	0.843103555	1.0000000000	-0.218493469	-0.115869590	-0.18615844
Wind	-0.004005673	0.05190883	-0.0026079544	-0.202012885	-0.244085773	-0.283195935	-0.2184934686	1.000000000	0.632193899	0.15217949
Gust	0.075943828	0.01482853	0.0072551349	-0.105404883	-0.156574670	-0.166324645	-0.1158695901	0.632193899	1.000000000	0.00677384
Pressure	-0.247442966	0.04137212	-0.0102542978	-0.205312616	0.355983313	0.102302239	-0.1861584413	0.152179492	0.006773840	1.00000000

```
> cor(subject$F[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.00000000	-0.19930337	0.064382136	0.19807314	-0.057675412	0.12883789	0.19053738	-0.04127626	-0.025407079	-0.04456497
HR	-0.19930337	1.00000000	0.002165860	-0.21948548	0.061126731	-0.13802071	-0.21553295	0.02357518	0.015579447	-0.01141282
Elevation	0.06438214	0.00216586	1.000000000	-0.01944663	-0.009380488	-0.02150494	-0.01890729	0.02172268	-0.002563425	-0.04688484
Temp	0.19807314	-0.21948548	-0.019446634	1.00000000	0.127011102	0.85864495	0.99777062	-0.15516609	-0.054550849	-0.28943873
Humidity	-0.05767541	0.06112673	-0.009380488	0.12701110	1.000000000	0.61050457	0.18899537	-0.39172517	-0.060137606	-0.21410775
DP	0.12883789	-0.13802071	-0.021504938	0.85864495	0.610504568	1.00000000	0.88833218	-0.29938188	-0.059903231	-0.34310844
HI	0.19053738	-0.21553295	-0.018907294	0.99777062	0.188995367	0.88833218	1.00000000	-0.17631116	-0.058198568	-0.30260882
Wind	-0.04127626	0.02357518	0.021722676	-0.15516609	-0.391725168	-0.29938188	-0.17631116	1.00000000	0.395726045	-0.11178794
Gust	-0.02540708	0.01557945	-0.002563425	-0.05455085	-0.060137606	-0.05990323	-0.05819857	0.39572604	1.000000000	-0.08847596
Pressure	-0.04456497	-0.01141282	-0.046884842	-0.28943873	-0.214107753	-0.34310844	-0.30260882	-0.11178794	-0.088475965	1.00000000

```
> cor(subject$G[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.000000000	-0.52640899	0.05511236	0.32372799	-0.12687077	0.21370174	0.3169083990	0.01792590	-0.0049489934	0.05334437
HR	-0.526408989	1.00000000	0.06121154	-0.14498072	0.04183537	-0.10010752	-0.1432452725	0.07611683	0.0601851906	0.03315095
Elevation	0.055112361	0.06121154	1.00000000	-0.01398286	-0.02525426	-0.03450604	-0.0154438150	0.01277720	0.0245049620	-0.03740521
Temp	0.323727992	-0.14498072	-0.01398286	1.00000000	-0.14936999	0.77756608	0.9982573191	0.02748191	0.0130921373	0.07265722
Humidity	-0.126870772	0.04183537	-0.02525426	-0.14936999	1.00000000	0.48867840	-0.0937575256	-0.19474002	-0.2175548427	0.37964708
DP	0.213701737	-0.10010752	-0.03450604	0.77756608	0.48867840	1.00000000	0.8103587199	-0.08527170	-0.1263768614	0.36329962
HI	0.316908399	-0.14324527	-0.01544382	0.99825732	-0.09375753	0.81035872	1.0000000000	0.01815133	0.0007766667	0.09325744
Wind	0.017925904	0.07611683	0.01277720	0.02748191	-0.19474002	-0.08527170	0.0181513274	1.00000000	0.4700756696	0.01404290
Gust	-0.004948993	0.06018519	0.02450496	0.01309214	-0.21755484	-0.12637686	0.0007766667	0.47007567	1.0000000000	-0.13738034
Pressure	0.053344373	0.03315095	-0.03740521	0.07265722	0.37964708	0.36329962	0.0932574373	0.01404290	-0.1373803362	1.00000000

```
> cor(subject$H[c(6:12, 15:17)])
```

	Pace	HR	Elevation	Temp	Humidity	DP	HI	Wind	Gust	Pressure
Pace	1.00000000	-0.75023694	0.02811084	0.04680125	-0.114776645	-0.02299990	0.03782727	0.14889454	0.047041136	0.08094489
HR	-0.75023694	1.00000000	-0.03357353	-0.14817685	0.155749024	-0.03502480	-0.13881158	-0.25319357	-0.092288128	-0.35845251
Elevation	0.02811084	-0.03357353	1.00000000	-0.01352037	-0.027508911	-0.02842386	-0.01608640	0.02229227	0.008389650	0.05218817
Temp	0.04680125	-0.14817685	-0.01352037	1.00000000	-0.128917969	0.77934318	0.99750057	0.06681749	0.051246427	-0.03416916
Humidity	-0.11477665	0.15574902	-0.02750891	-0.12891797	1.000000000	0.51403816	-0.06086407	-0.21975143	-0.007456559	-0.26599896
DP	-0.02299990	-0.03502480	-0.02842386	0.77934318	0.514038159	1.00000000	0.81956378	-0.05463759	0.055936817	-0.18869943
HI	0.03782727	-0.13881158	-0.01608640	0.99750057	-0.060864074	0.81956378	1.00000000	0.05201118	0.051238193	-0.05092578
Wind	0.14889454	-0.25319357	0.02229227	0.06681749	-0.219751434	-0.05463759	0.05201118	1.00000000	0.625110177	0.22852809
Gust	0.04704114	-0.09228813	0.00838965	0.05124643	-0.007456559	0.05593682	0.05123819	0.62511018	1.000000000	0.04830350
Pressure	0.08094489	-0.35845251	0.05218817	-0.03416916	-0.265998960	-0.18869943	-0.05092578	0.22852809	0.048303505	1.00000000

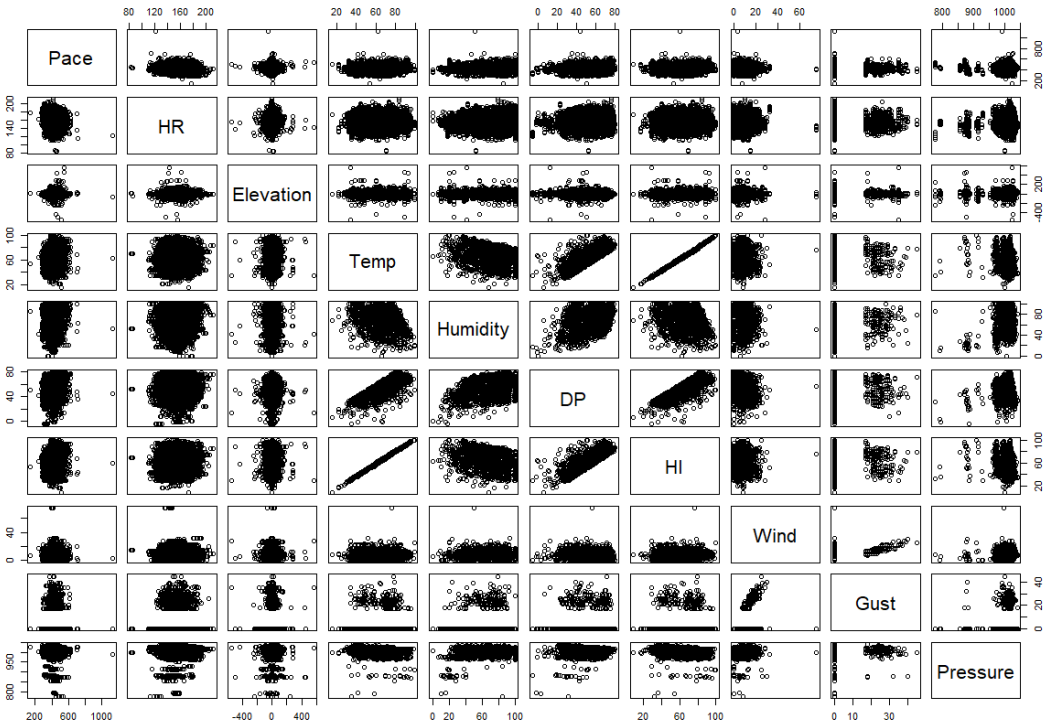
On reviewing the correlation tables for each individual data subject it is apparent that there are better rates of correlation available here depending on the specific subject. This can likely be attributed to the idiosyncrasies relating to each subject's training style, be it typical distance covered, effort level, time of day they conduct their workout etc.

There is one stand out subject from each gender, subject C (female) and G (male). Subject H (male) also offers some interesting correlations.

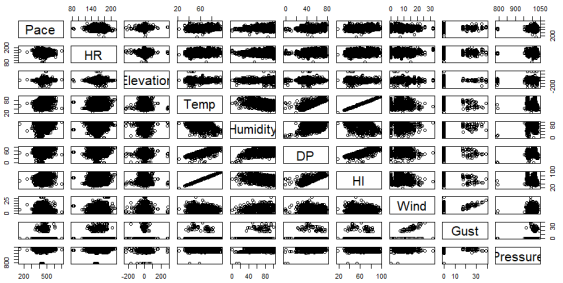
Correlation Matrices

The graphs presented below visualise the numeric data presented above and so little further interpretation is required. Correlation matrices are shown for each of the test criteria above, full dataset, each gender and then each data subject.

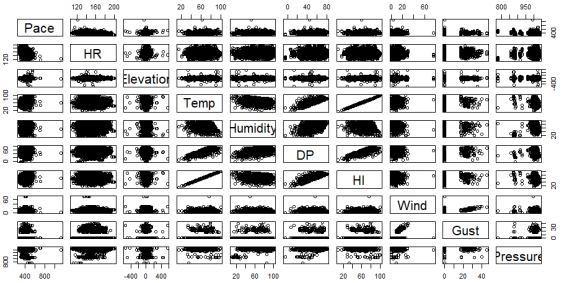
```
> pairs(dataset[c(6:12, 15:17)])
```



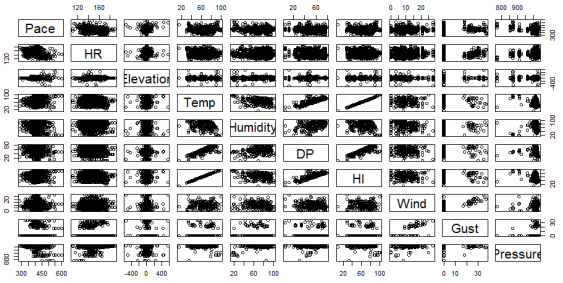
```
> pairs(gender$F[c(6:12, 15:17)])
```



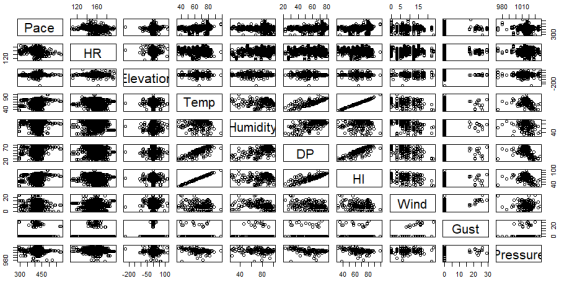

```
> pairs(gender$M[c(6:12, 15:17)])
```



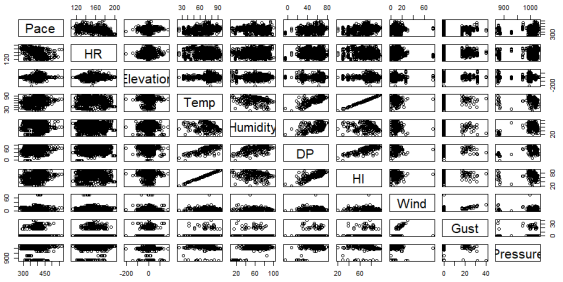

```
> pairs(subject$A[c(6:12, 15:17)])
```



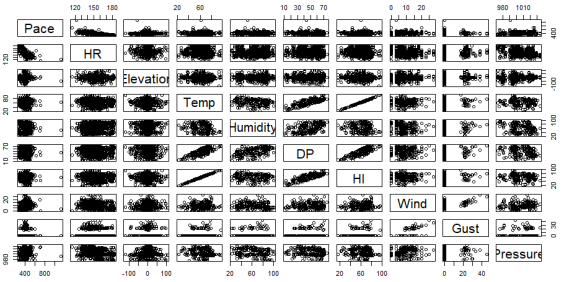

```
> pairs(subject$B[c(6:12, 15:17)])
```



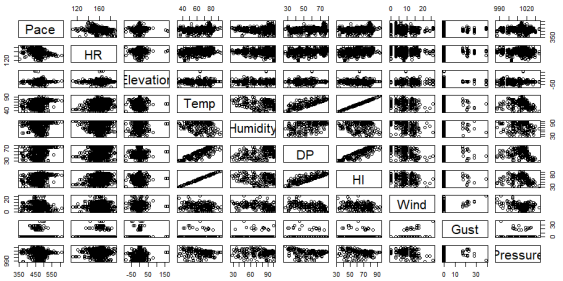

```
> pairs(subject$C[c(6:12, 15:17)])
```



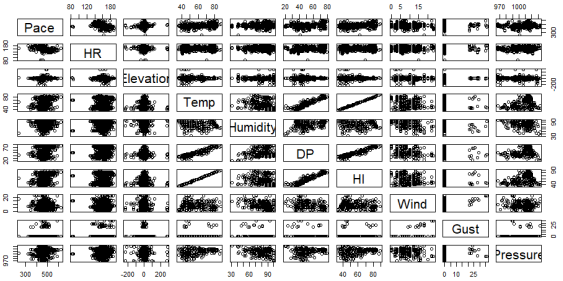

```
> pairs(subject$D[c(6:12, 15:17)])
```



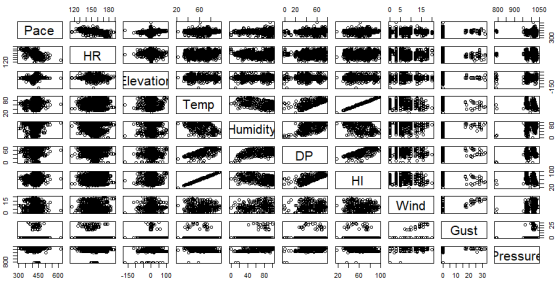

```
> pairs(subject$E[c(6:12, 15:17)])
```



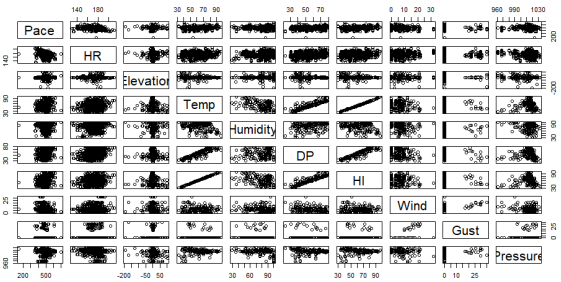

```
> pairs(subject$F[c(6:12, 15:17)])
```



```
> pairs(subject$G[c(6:12, 15:17)])
```




```
> pairs(subject$H[c(6:12, 15:17)])
```



The reason for the apparent lack of correlation between the expected variables of HR and Pace in respect to Temp, DP and HI could be subtle. The correlation values of between 0.2 and 0.1 could be disguising a minimal cause-effect relationship that exists between the two which is further obscured by the relatively minimal number of data observations. A much larger dataset might weed out a lot more of the variability that exists within metrics of this type. A larger dataset could of course also just exacerbate the situation.

Summary Statistics

```
> summary(dataset)
```

```
     Name         Gender        Date               Time             Location            Pace      
Adam   :1421   F:3382   Length:7922        Length:7922        Length:7922        Min.   :144  
Matt   :1418   M:4540   Class :character   Class :character   Class :character   1st Qu.:400  
Alex   :1307            Mode  :character   Mode  :character   Mode  :character   Median :422  
Grant  : 869                                                                     Mean   :426  
Joe    : 832                                                                     3rd Qu.:453  
Sally  : 714                                                                     Max.   :716  
Molly  : 692                                                                                  
Jordan : 669                                                                                  
       HR          Elevation             Temp          Humidity          DP              HI       
 Min.   : 83.0   Min.   :-551.0000   Min.   :15.00   Min.   :  0.00   Min.   :-5.00   Min.   :  9.0  
 1st Qu.:145.0   1st Qu.:  -3.0000   1st Qu.:53.25   1st Qu.: 58.00   1st Qu.:42.00   1st Qu.: 52.0  
 Median :155.0   Median :   0.0000   Median :67.00   Median : 76.00   Median :56.00   Median : 67.0  
 Mean   :155.4   Mean   :   0.1004   Mean   :65.13   Mean   : 71.23   Mean   :54.14   Mean   : 64.7  
 3rd Qu.:165.0   3rd Qu.:   4.0000   3rd Qu.:77.00   3rd Qu.: 88.00   3rd Qu.:68.00   3rd Qu.: 78.0  
 Max.   :211.0   Max.   : 564.0000   Max.   :99.00   Max.   :100.00   Max.   :80.00   Max.   :100.0  
          DewFeel        RealFeel         Wind             Gust          Pressure     
 Comfortable  : 707   Caution:1336   Min.   : 0.000   Min.   : 0.000   Min.   : 775  
 Dry          :3020   Okay   :6586   1st Qu.: 5.000   1st Qu.: 0.000   1st Qu.: 998  
 Miserable    : 295                  Median : 7.000   Median : 0.000   Median :1013  
 Oppressive   :1079                  Mean   : 8.021   Mean   : 2.496   Mean   :1005  
 Pleasant     : 918                  3rd Qu.:12.000   3rd Qu.: 0.000   3rd Qu.:1018  
 Sticky       : 907                  Max.   :75.000   Max.   :45.000   Max.   :1040  
 Uncomfortable: 996                                                                  
```

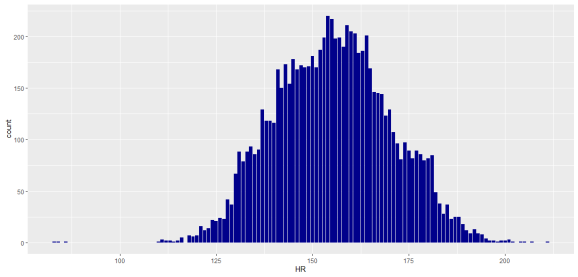
The first noteworthy observation from this dataset summary is the apparently large number of zero values in the Wind and Gust variables. This is supported in viewing the correlation graphs above. This observation might prove useful during later testing where it is likely that these variables will play little role in the prediction of response variables.

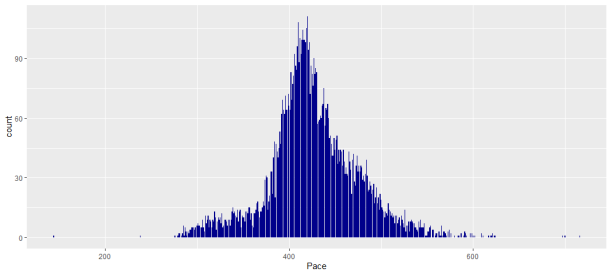
No other variables offer any insight outside of what might be expected. The only other point of note is that the median and mean values for Elevation are both essentially zero, suggesting that the subjects sampled tend to run on fairly flat terrain. A point which is supported by the fact that the primary regions sampled in, Southern Louisiana and Southeast Texas, are widely considered to be areas of largely flat terrain.

Barplots

The HR and Pace attributes are visualised below in order to understand if these variables present in a normally distributed fashion.

```
> ggplot(dataset, aes(HR)) + geom_bar(fill = "dark blue")
> ggplot(dataset, aes(Pace)) + geom_bar(fill = "dark blue")
```






```
> shapiro.test(table(dataset$HR))

    Shapiro-Wilk normality test

    data:  table(dataset$HR)
    W = 0.86717, p-value = 6.073e-08
```

```
> shapiro.test(table(dataset$Pace))

    Shapiro-Wilk normality test

    data:  table(dataset$Pace)
    W = 0.79823, p-value < 2.2e-16
```

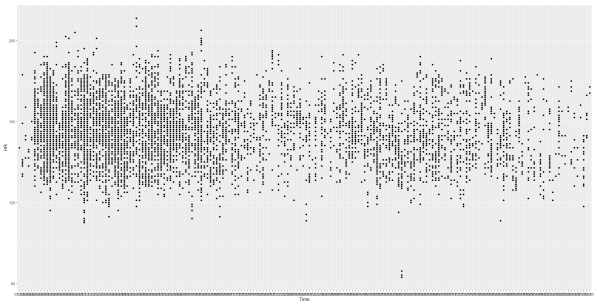
While both graphs visually appear to be normally distributed at first sight, the HR graph more so than the Pace graph, the shapiro-wilk test for normality is used to confirm this assumption.

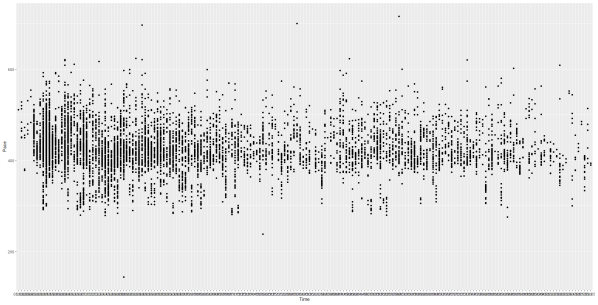
The null-hypothesis for both tests states that the population is normally distributed for the relevant attribute. Each of the p-values presented from these tests show a certainty level well below the 0.01 alpha. Therefore the null hypothesis can be rejected suggesting that there is evidence the data tested are not normally distributed. This could be explained by outliers within the data, which could be supported by the evidence of small bars to the outer left and right regions of the graphs above.

Scatterplots

Scatterplots were used to understand if there was any connection between the Time at which an activity was conducted and higher or lower instances of the running based metrics HR and Pace.

```
> p1 <- ggplot(data = dataset, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> p2 <- ggplot(data = dataset, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```

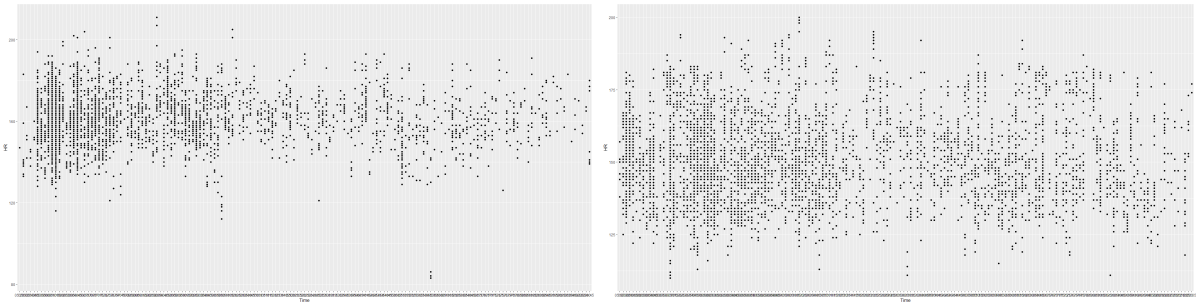




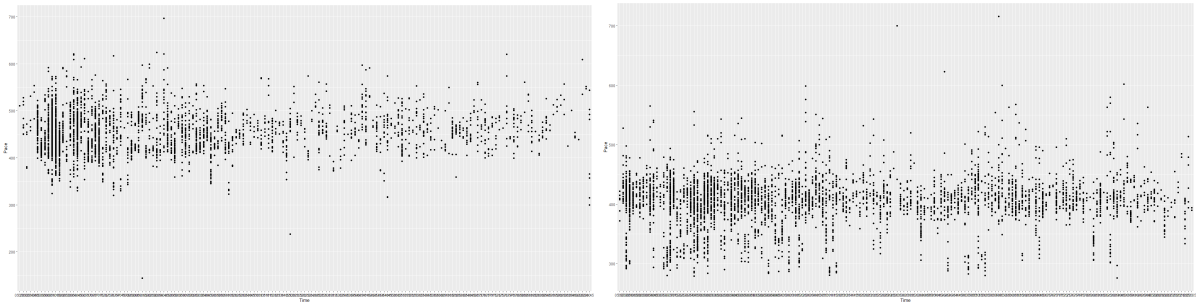
Neither plot suggests that the data overall conforms to any such assumptions that the hour at which a workout is conducted has any significant effect on the measured physical output (slower paces or faster heart rates). It remains pretty uniform throughout with an apparent greater density of records in earlier hours. (This was expected as mentioned earlier, athletes tend to avoid the daily extremes especially when residing in difficult environments.)

As seen below the plots don't appear to offer any further insight even when broken down further by gender or individual data subject.

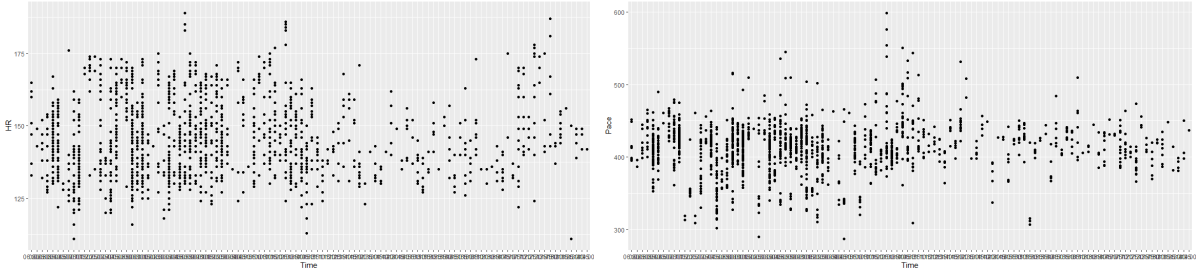
```
> g1 <- ggplot(data = gender$F, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> g3 <- ggplot(data = gender$M, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



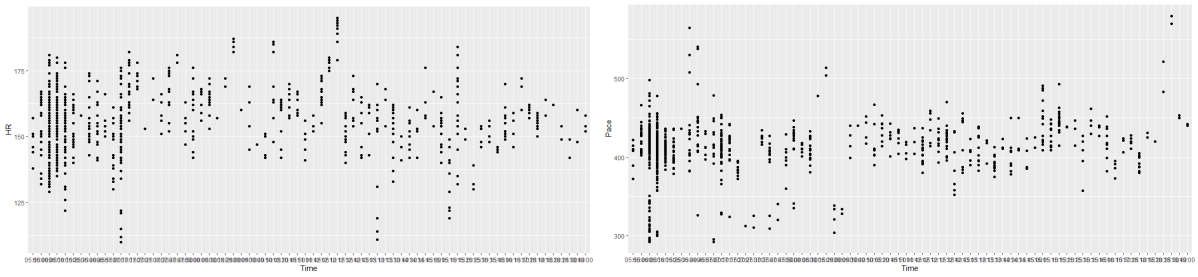
```
> g2 <- ggplot(data = gender$F, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> g4 <- ggplot(data = gender$M, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



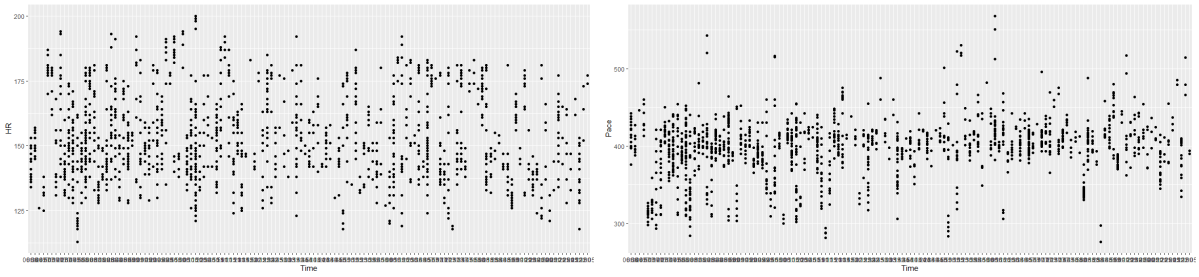
```
> s1 <- ggplot(data = subject$E, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s2 <- ggplot(data = subject$E, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



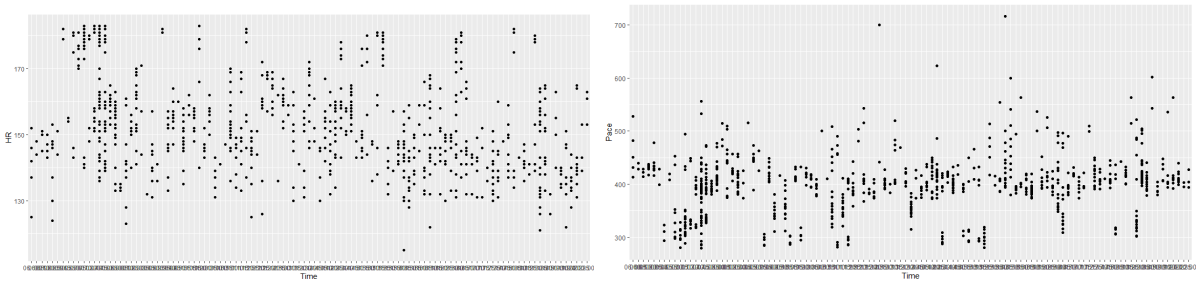
```
> s3 <- ggplot(data = subject$F, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s4 <- ggplot(data = subject$F, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



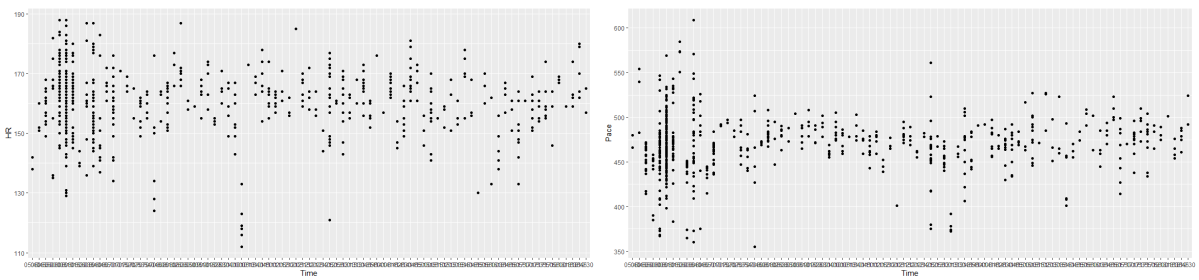

```
> s5 <- ggplot(data = subject$G, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s6 <- ggplot(data = subject$G, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



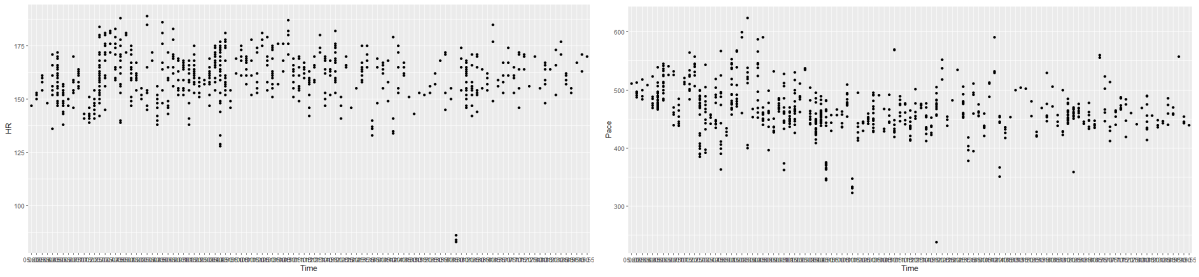

```
> s7 <- ggplot(data = subject$H, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s8 <- ggplot(data = subject$H, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



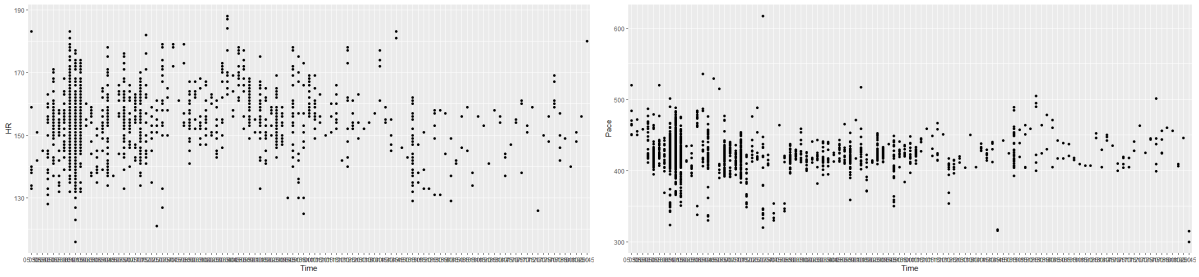

```
> s9  <- ggplot(data = subject$A, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s10 <- ggplot(data = subject$A, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



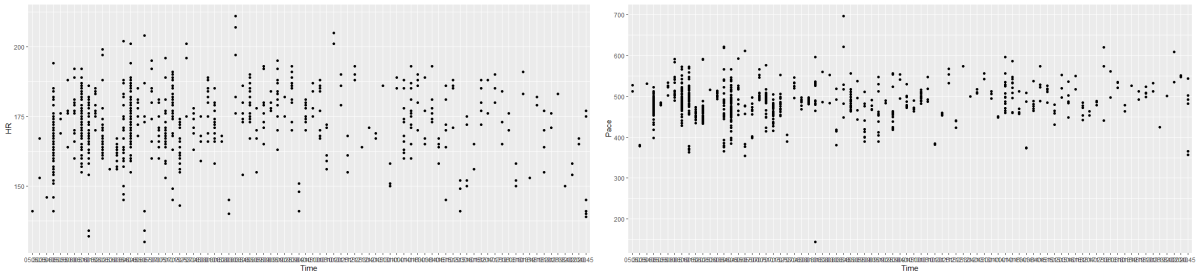
```
> s11 <- ggplot(data = subject$B, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s12 <- ggplot(data = subject$B, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```




```
> s13 <- ggplot(data = subject$C, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s14 <- ggplot(data = subject$C, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



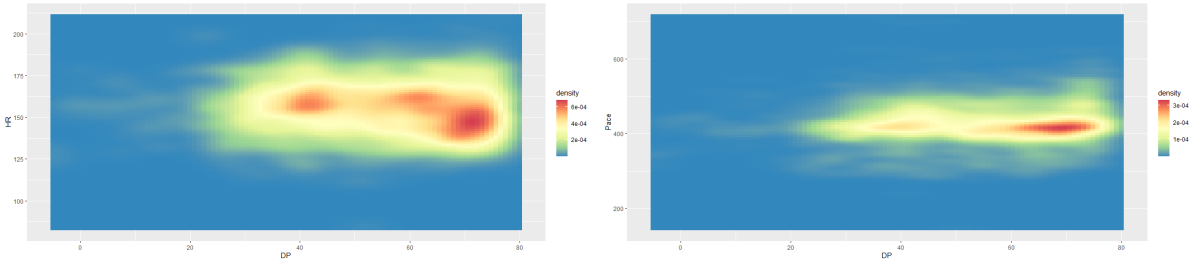

```
> s15 <- ggplot(data = subject$D, aes(Time, HR)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
> s16 <- ggplot(data = subject$D, aes(Time, Pace)) + geom_point()
    + stat_smooth(method = "gam", formula = y ~ s(x), size = 1)
```



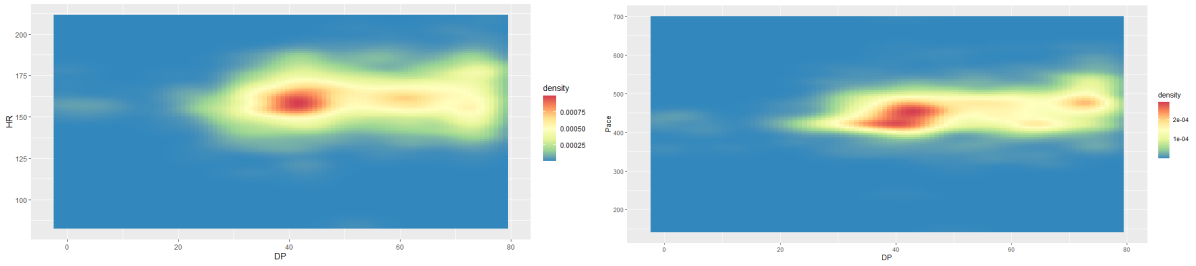
Density Graphs

The density graphs produced below offer an insight into the most common HR and Pace values in respect to the DP values measured during activities. These graphs should be able to support the understanding that the dataset is made up of individuals that have extensive experience in humid environments as evidenced by the high DP values.

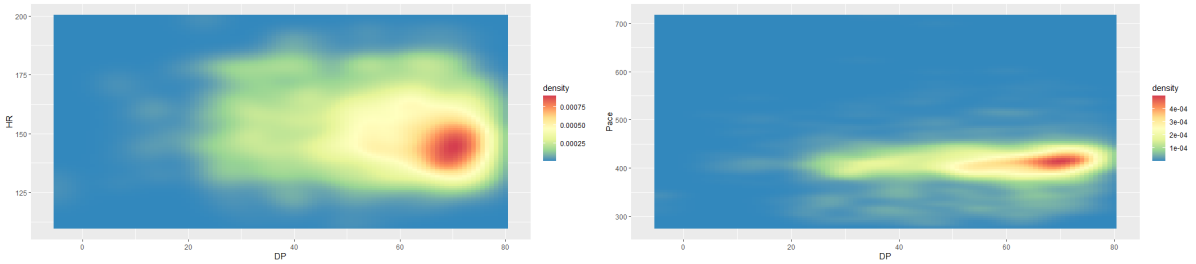
```
> ggplot(dataset, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(dataset, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



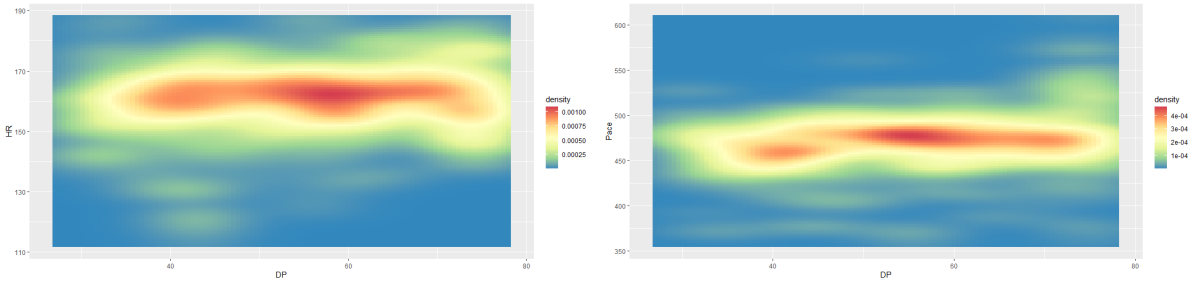

```
> ggplot(gender$F, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(gender$F, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



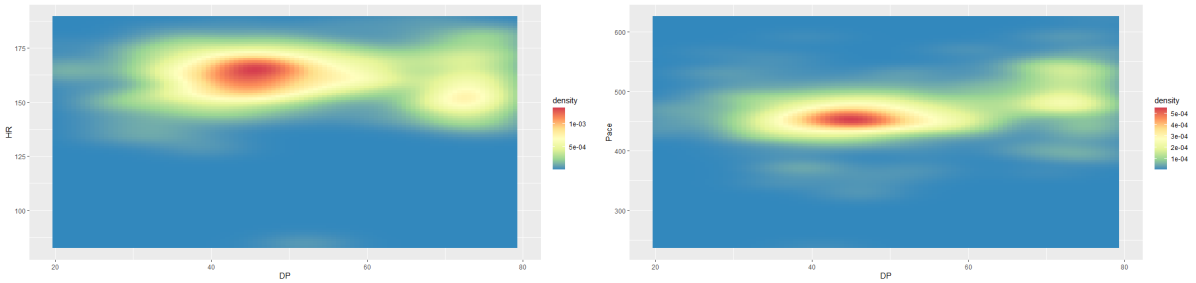

```
> ggplot(gender$M, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(gender$M, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



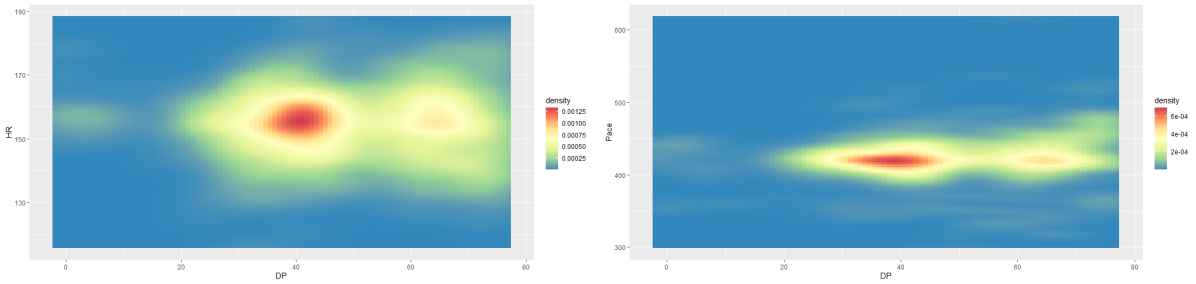
```
> ggplot(subject$A, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$A, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



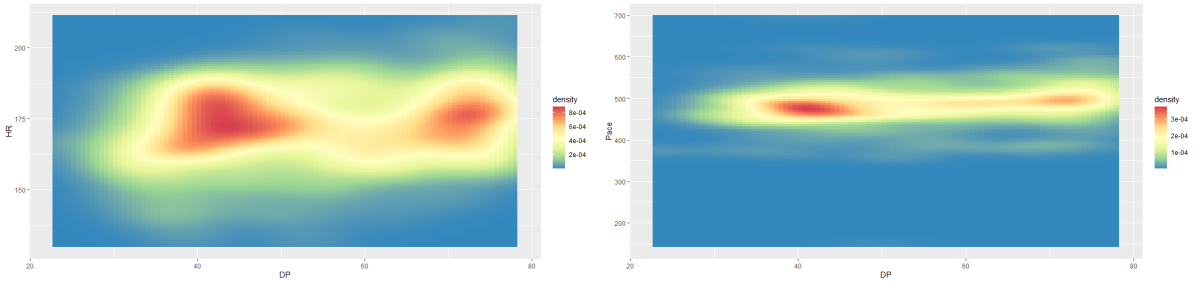

```
> ggplot(subject$B, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$B, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



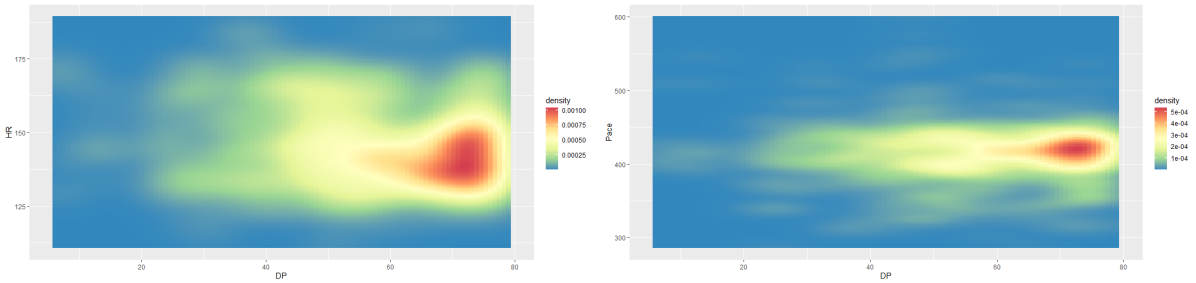

```
> ggplot(subject$C, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$C, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



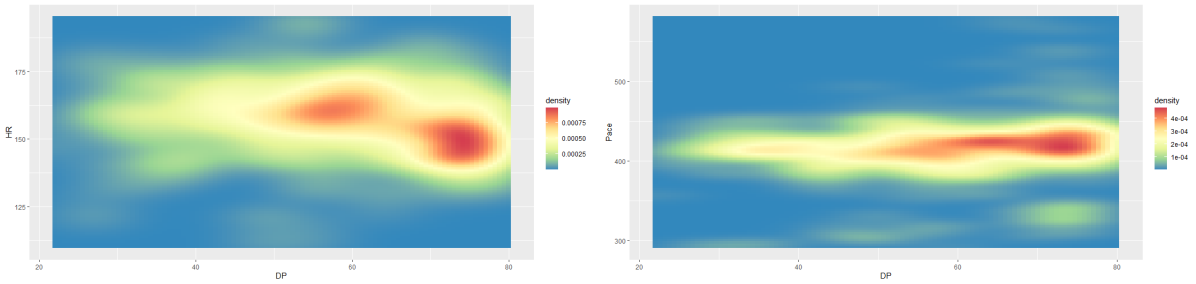
```
> ggplot(subject$D, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$D, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



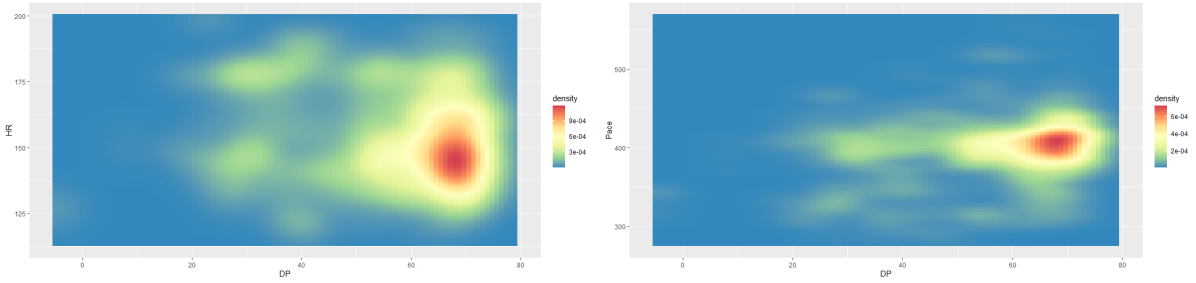

```
> ggplot(subject$E, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$E, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



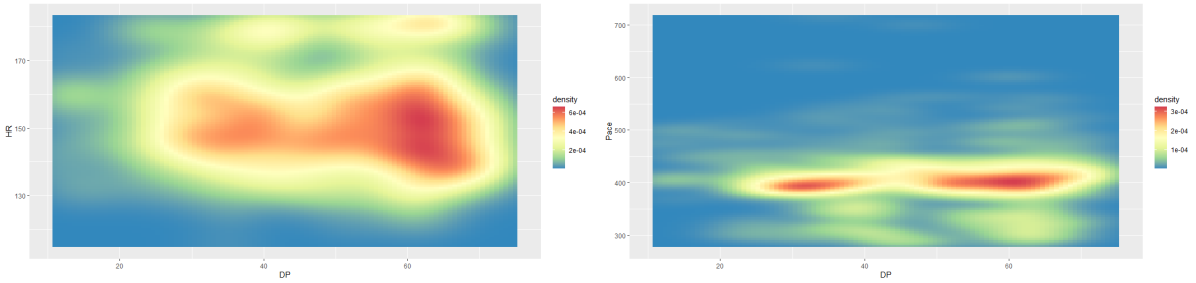

```
> ggplot(subject$F, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$F, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```



```
> ggplot(subject$G, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$G, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```




```
> ggplot(subject$H, aes(DP, HR))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')

> ggplot(subject$H, aes(DP, Pace))
+ stat_density2d(geom="tile", contour=FALSE, aes(fill =..density..))
+ scale_fill_distiller(palette = 'Spectral')
```

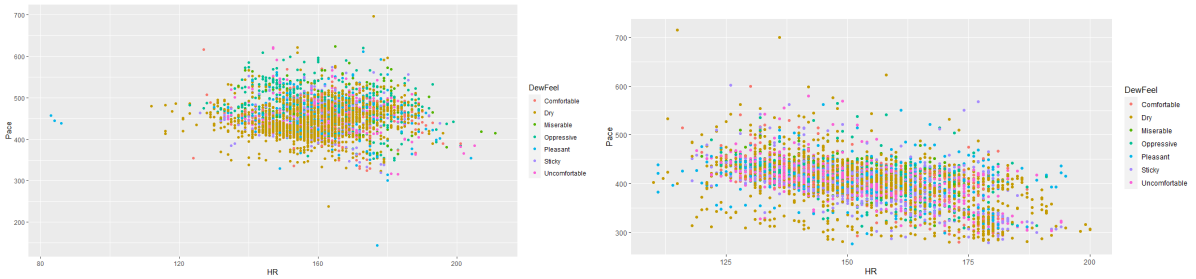


Clustering

The HR and Pace variables are plotted against each other and clustered based on the DewFeel categories. This visualisation offers an insight into whether or not these specific humidity ranges have any effect on the performance metrics.

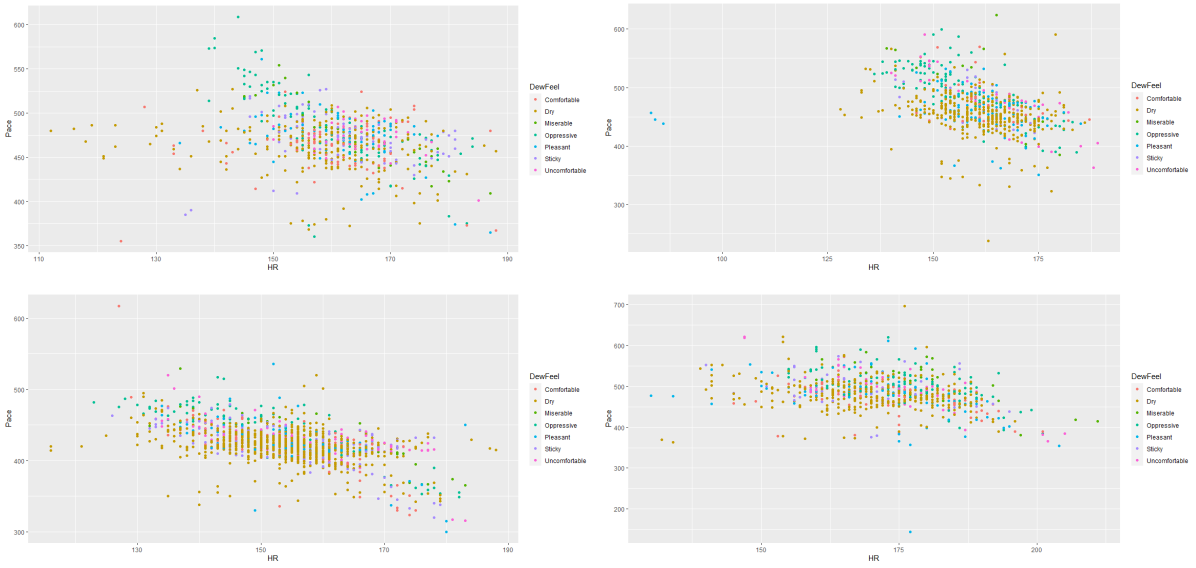
```
> ggplot(dataset, aes(HR, Pace, colour = DewFeel)) + geom_point()
```

```
> ggplot(gender$F, aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot(gender$M, aes(HR, Pace, colour = DewFeel)) + geom_point()
```

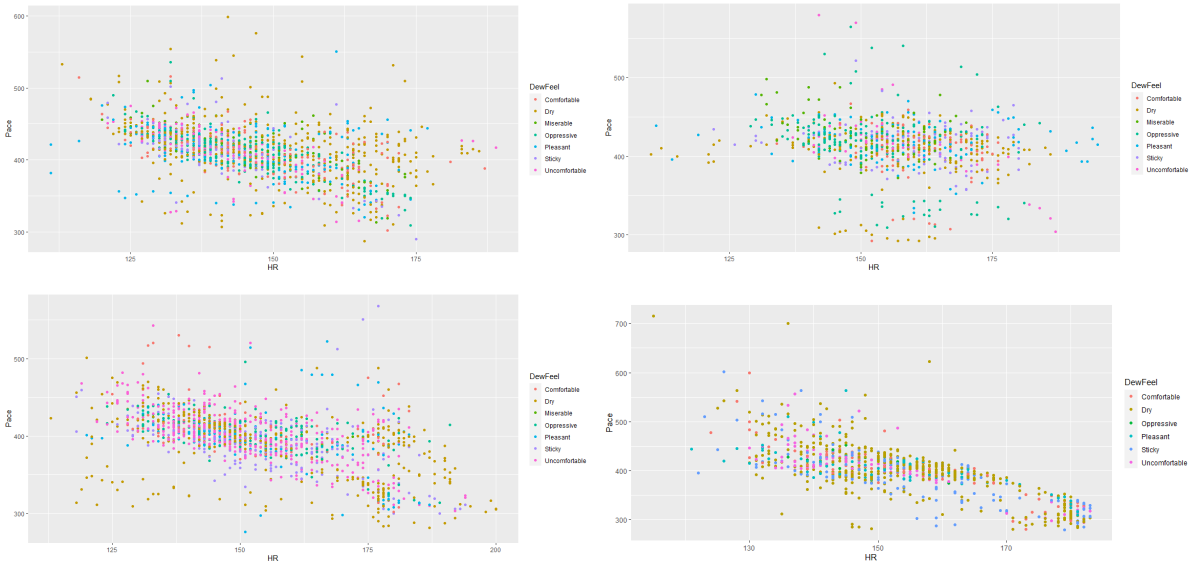


These figures make it clear to the observer that Dew Point (DewFeel) categories (outlined in the appendix) do not appear to show any clustering behaviour in respect to HR and Pace variables. The same is true when considering output from each individual data subject.

```
> ggplot((subject$A), aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot((subject$B), aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot((subject$C), aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot((subject$D), aes(HR, Pace, colour = DewFeel)) + geom_point()
```



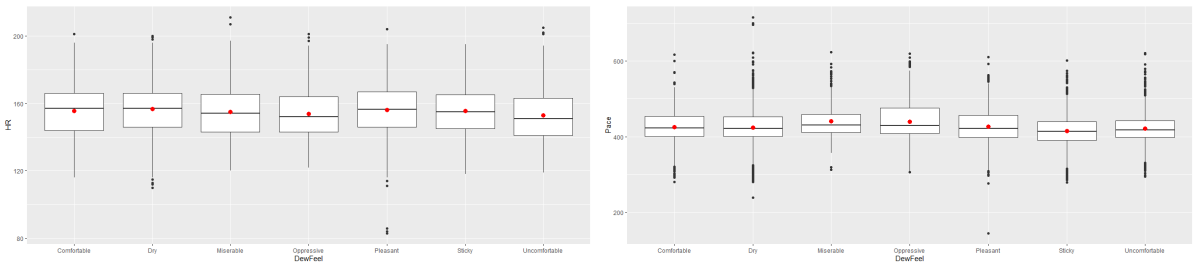
```
> ggplot((subject$E), aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot((subject$F), aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot((subject$G), aes(HR, Pace, colour = DewFeel)) + geom_point()
> ggplot((subject$H), aes(HR, Pace, colour = DewFeel)) + geom_point()
```



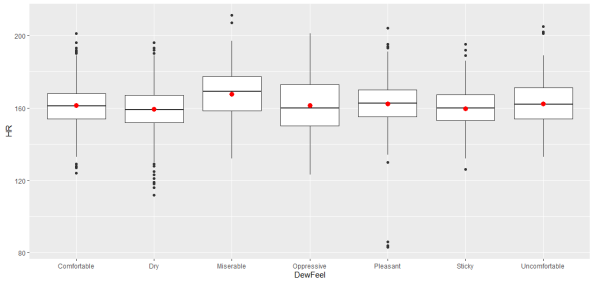
Box Plots

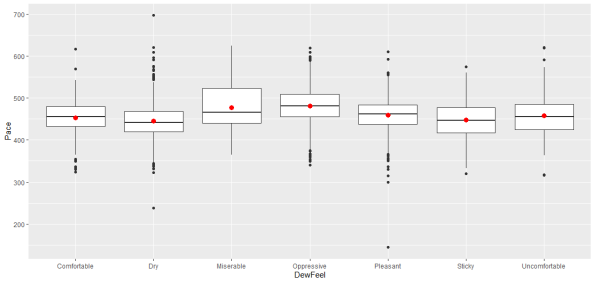
The boxplots below show an apparent lack of variability of HR and Pace between each of the DewFeel categories. The subtle differences seem to fall in line with the rest of the analysis produced in this report, suggesting that it may be difficult to prove or predict HR and Pace as response variables.

```
> ggplot(dataset, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(dataset, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```

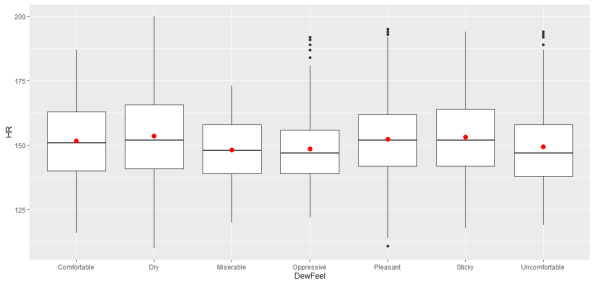


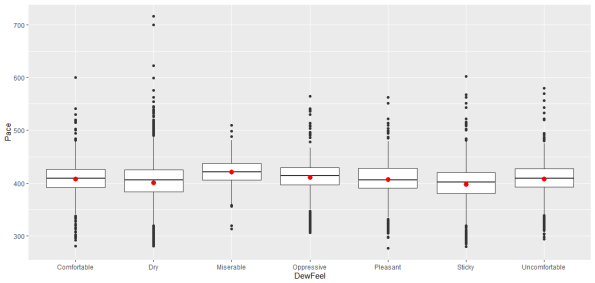
```
> ggplot(gender$F, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(gender$F, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```



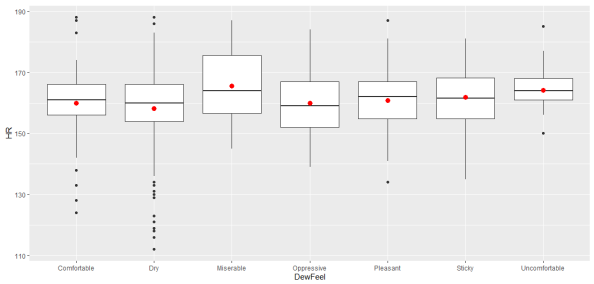


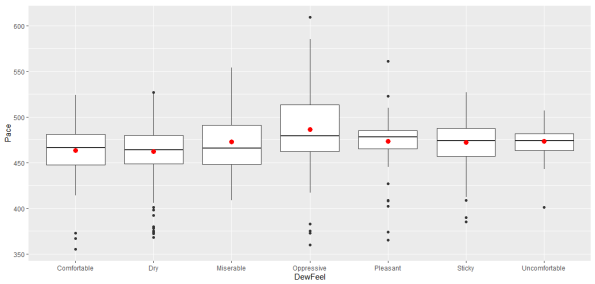

```
> ggplot(gender$M, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(gender$M, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```



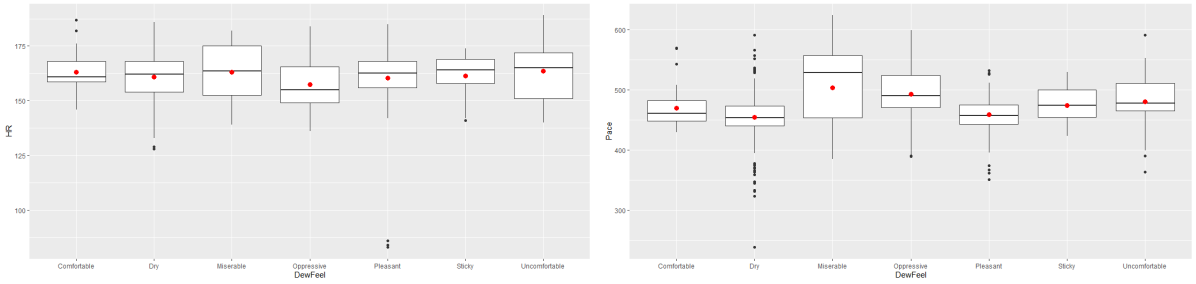



```
> ggplot(subject$A, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$A, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```

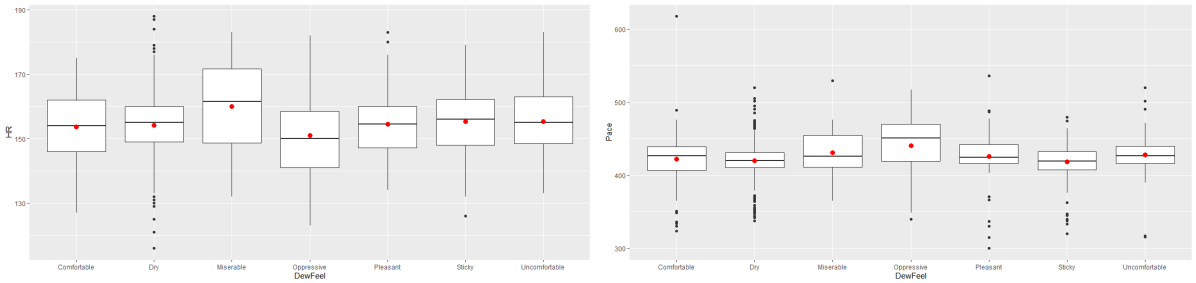




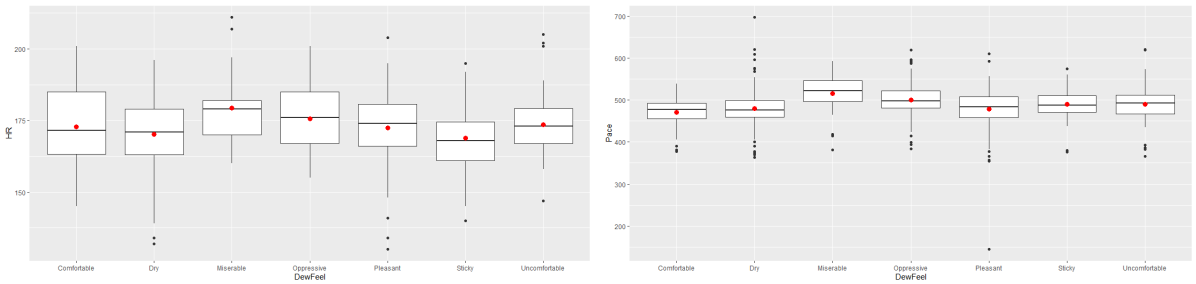
```
> ggplot(subject$B, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$B, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```



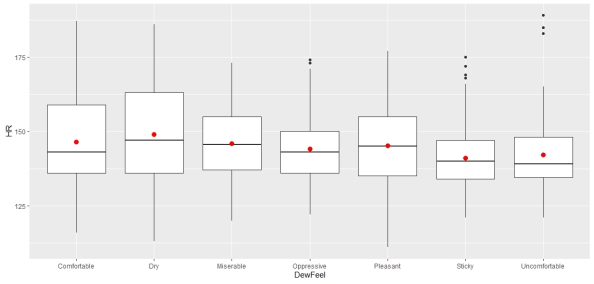

```
> ggplot(subject$C, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$C, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```

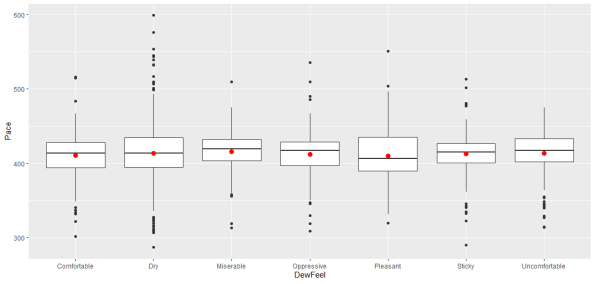



```
> ggplot(subject$D, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$D, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```

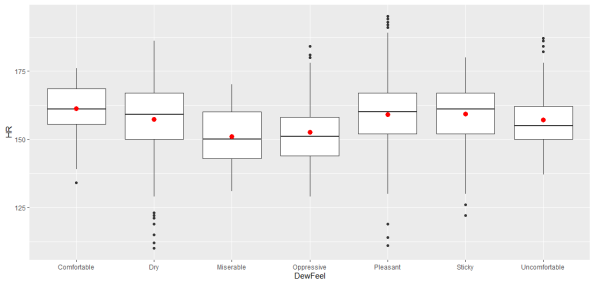


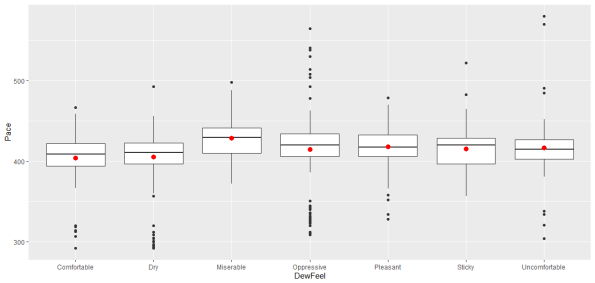
```
> ggplot(subject$E, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$E, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```



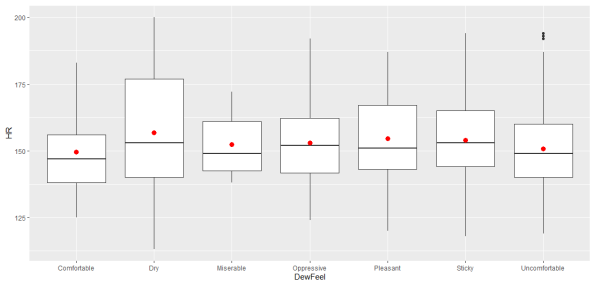


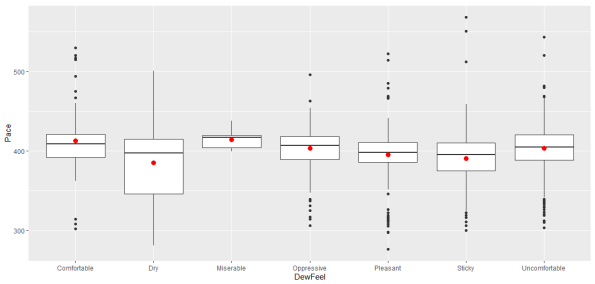

```
> ggplot(subject$F, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$F, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```



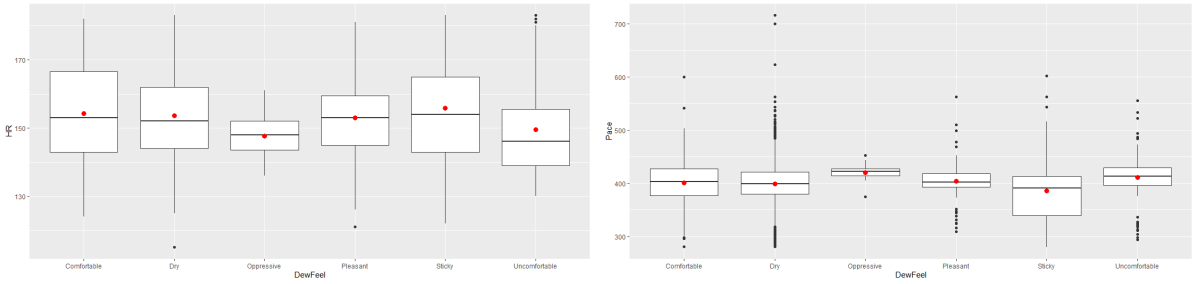



```
> ggplot(subject$G, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$G, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```





```
> ggplot(subject$H, aes(DewFeel, HR))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)

> ggplot(subject$H, aes(DewFeel, Pace))+ geom_boxplot()
+ stat_summary(fun = mean, geom = "point", color = "red", size = 3)
```



Categorical Variables

Contingency Table(s)

The contingency table presented allows the observer to understand numerically the consistency with which each data subject was subjected to the differing levels of dew point humidity. The table shows a healthy mix of data points in each of the three 'humid' categories as well as the three 'non-humid' categories for each data subject.

```
> contingency <- table(dataset$Name, dataset$DewFeel)
```

	Comfortable	Dry	Miserable	Oppressive	Pleasant	Sticky	Uncomfortable
A	88	242	31	106	80	72	50
B	40	345	14	123	100	29	41
C	86	714	30	107	82	164	124
D	62	297	29	117	90	55	64
E	157	423	106	298	184	99	151
F	87	213	78	161	126	106	61
G	97	365	7	156	161	221	414
H	90	421	0	11	95	161	91

The number of occurrences of a Dew Point value within a category is dependent upon the data subject due to the way in which the data was collected and the lack of overlap between these athletes running at the same time in the same location. A chi-square test can be used to prove this dependent relationship statistically.

H_0: The categorical variables Name and DewFeel are independent

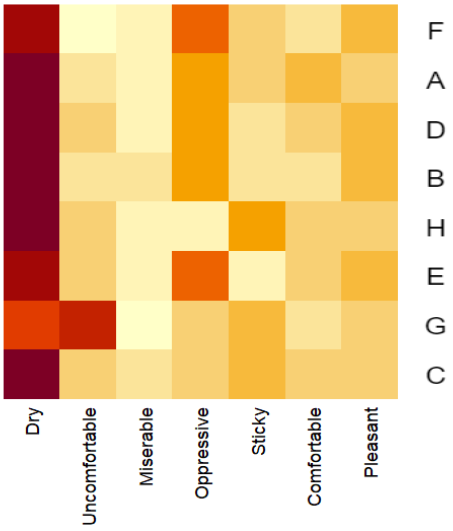
H_1: The categorical variables Name and DewFeel are dependent

```
> chisq.test(contingency)

        Pearson's Chi-squared test

data:  contingency
X-squared = 1314.6, df = 42, p-value < 2.2e-16
```

The small p-value observed allows for the rejection of the null hypothesis, suggesting that the variables in this contingency table are dependent upon each other at a significance level far less than 0.01.

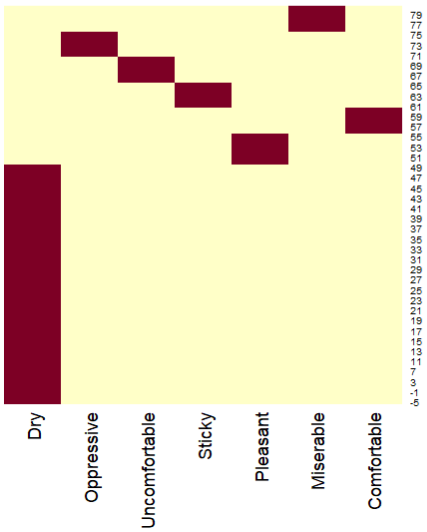
Heatmap(s)

```
> heatmap(contingency)
```



While the large majority of the occurrences for each data subject in this map are evidently coming from the 'Dry' category, this occurrence can be explained by the method in which the categories were calculated. The dry category is simply set as all DP degrees below 50℉ (see appendix). It is still evident to see from the heatmap that each individual has had some viable time and experience in some of the much more difficult DewFeel categories which are all only separated by 5℉ of range, with the exception of 'Miserable, category which has no ceiling value, similar to the 'Dry' category without a floor. In this instance however it is much more difficult for the environment to even reach DP temperatures in that category.

This heatmap of DewFeel versus DP is shown for context in order to aid in understanding the differing nature of the Dry category.




```
          DewFeel
Comfortable  : 707
Dry          :3020
Miserable    : 295
Oppressive   :1079
Pleasant     : 918
Sticky       : 907
Uncomfortable: 996
```

It is immediately evident in this visualisation of data, which was presented numerically in the summary function above (also shown here for reference), that the 'Dry' category of DewFeel covers a far greater range of DP temperatures and so overwhelms the data.

Simple Linear Regression

Hypothesis:

A change in the Dew Point (DP) will cause a similar change in Heart Rate (HR) values.

(This will be considered the alternate hypothesis, the null hypothesis states that there is no cause-effect relationship between HR and DP).

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

```
> summary(lm(HR ~ DP, dataset))

Call:
lm(formula = HR ~ DP, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-72.542 -10.800  -0.219  10.090  56.707

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 158.24970    0.60889 259.898  < 2e-16 ***
DP           -0.05206    0.01081  -4.817 1.48e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.98 on 7920 degrees of freedom
Multiple R-squared:  0.002921,  Adjusted R-squared:  0.002796
F-statistic: 23.21 on 1 and 7920 DF,  p-value: 1.483e-06
```

The output shows an average HR intercept at 158.25 beats per minute (bpm). The regression model then predicts that with a rise in DP that there will be a decrease in the HR value. The exact formula, and therefore rate of decrease with a rise in the DP value, is shown:

Response = Intercept + (β ∗ Predictor)
HR = 158.249 + (-0.052 * DP)

The model shows significance well below the more stringent p-value alpha of 0.01. Allowing for the rejection of the null hypothesis which would give credence to the alternate hypothesis set out above. What is interesting is that the regression is suggesting that the change in HR is occurring in the opposite direction than it was expected, with declining HR in response to a rising DP, rather than an increasing HR. A possible reason for this could be that there is a reduction in the Pace at which athletes are running, when the DP values are higher and it is more difficult to perform, resulting in a decrease in HR rather than an increase. Further analysis should shed light on this case.

While there is p-value significance present that could lead to the rejection of the null hypothesis. β_1 is only ≈ 0.05 which is miniscule when considered in conjunction with this particular data. It would require an increase of 20℉ in the DP just to predict a decrease of a single beat per minute in HR. The β_1 is also incredibly close to zero, which itself would actually fail to reject the null hypothesis. Therefore the significance of this model could be called into question.

Hypothesis: A change in the DP will result in change in Pace values.

(This will be considered the alternate hypothesis, the null hypothesis will be the inverse)

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

```
> summary(lm(Pace ~ DP, dataset))

Call:
lm(formula = Pace ~ DP, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-280.872  -25.738   -3.566   27.679  295.261

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 411.09463    1.97432 208.221  < 2e-16 ***
DP            0.27555    0.03504   7.863 4.25e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48.57 on 7920 degrees of freedom
Multiple R-squared:  0.007746,      Adjusted R-squared:  0.00762
F-statistic: 61.83 on 1 and 7920 DF,  p-value: 4.248e-15
```

It was suspected both before analysis and from the previous regression analysis that Pace values might increase, that is to say that mile segments are run slower, in response to increasing DP values due to the more difficult environmental conditions caused by higher DP values. The above model suggests exactly that, the regression model is predicting that Pace increases at a rate of one second per mile with an approximately 4℉ increase in the DP. The formula calculated is shown here:

Response = Intercept + (β ∗ Predictor)
Pace = 411.095 + (0.27555 * DP)

It is important to acknowledge that the p-values calculated are well below the lower 0.01 level of significance which again allows for the rejection of the null hypothesis. Offering support for the alternate hypothesis proposed, that increasing DP values result in increasing Pace values. The $\beta_1 \approx 0.28$ in this case also gives much more credence to this calculation than the previous HR calculation as it has a much more measurable effect on the values even if these changes may be impreceivable to the data subjects themselves.

The same hypothesis is posed for the response variables of HR and Pace, but this time against the predictor values of Heat Index (HI). This is just to assess if there is a similar relationship caused by a different weather predictor metric rather than just the humidity based DP.

```
> summary(lm(HR ~ HI, dataset))

Call:
lm(formula = HR ~ HI, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-71.924 -10.625  -0.215   9.845  57.965

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 163.0672     0.7057  231.06   <2e-16 ***
HI           -0.1180     0.0106  -11.14   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.89 on 7920 degrees of freedom
Multiple R-squared:  0.01542, Adjusted R-squared:  0.0153
F-statistic:   124 on 1 and 7920 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ HI, dataset))

Call:
lm(formula = Pace ~ HI, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-281.021  -25.820   -4.311   27.465  291.742

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 420.52535    2.31093 181.972   <2e-16 ***
HI            0.08483    0.03470   2.445   0.0145 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.74 on 7920 degrees of freedom
Multiple R-squared:  0.000754,Adjusted R-squared:  0.0006278
F-statistic: 5.976 on 1 and 7920 DF,  p-value: 0.01452
```

Both models show significance in their calculations, with HR ~ HI generating a p-value far below the 0.01 threshold, and Pace ~ HI just at that threshold with a p-value of 0.0145.

The HR ~ HI model appears to be much more significant in its predictions than the HR ~ DP model. This time predicting a decrease in HR for every 10℉ increase in HI temperature. This is essentially a category change in the RealFeel variable and offers support for the previous model that also predicted a decreasing HR with increasing weather metrics. The formula calculated by the model is shown here:

Response = Intercept + (β * Predictor)
HR = 163.067 + (-0.118 * HI)

The Pace ~ HI on the other hand is far less significant in its predictions than the previous model of Pace ~ DP. On this occasion requiring an approximate increase of 12℉ in the HI measurement in order to cause a single second increase in the Pace recorded. The $\beta_1 \approx 0.085$, which isn't terribly close to the zero value that would fail to reject the null hypothesis but it is close enough to mention that trust in this model could be called into question. The formula generated by the model shown here:

Response = Intercept + (β * Predictor)
HR = 163.067 + (-0.118 * HI)

Similar models are applied to the male and female subsets as well as the individual subsets using DP as a predictor for the HR and Pace performance metrics, in order to determine if similar calculations can be generated for calculating HR/Pace in response to a changing DP.

```
> summary(lm(HR ~ DP, gender$F))

Call:
lm(formula = HR ~ DP, data = gender$F)

Residuals:
    Min      1Q  Median      3Q     Max
-77.515  -7.785  -0.202   8.073  48.131

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 155.41518    0.79000 196.729  < 2e-16 ***
DP            0.09808    0.01454   6.746 1.78e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.96 on 3380 degrees of freedom
Multiple R-squared:  0.01328, Adjusted R-squared:  0.01299
F-statistic:  45.5 on 1 and 3380 DF,  p-value: 1.784e-11
```

```
> summary(lm(HR ~ DP, gender$M))

Call:
lm(formula = HR ~ DP, data = gender$M)

Residuals:
    Min      1Q  Median      3Q     Max
-42.341 -11.171  -1.267  10.404  44.891

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 157.55742    0.83362 189.003  < 2e-16 ***
DP           -0.10645    0.01443  -7.379 1.89e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.15 on 4538 degrees of freedom
Multiple R-squared:  0.01186, Adjusted R-squared:  0.01164
F-statistic: 54.45 on 1 and 4538 DF,  p-value: 1.885e-13
```

```
> summary(lm(HR ~ DP, subject$A))

Call:
lm(formula = HR ~ DP, data = subject$Jordan)

Residuals:
    Min      1Q  Median      3Q     Max
-46.726  -5.819   0.948   6.809  30.321

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 153.72357    1.79706  85.542  < 2e-16 ***
DP            0.11632    0.03145   3.698 0.000235 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.22 on 667 degrees of freedom
Multiple R-squared:  0.0201,  Adjusted R-squared:  0.01863
F-statistic: 13.68 on 1 and 667 DF,  p-value: 0.0002347
```

```
> summary(lm(HR ~ DP, subject$B))

Call:
lm(formula = HR ~ DP, data = subject$Molly)

Residuals:
    Min      1Q  Median      3Q     Max
-77.590  -6.699   0.494   7.373  28.559

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 161.105535   1.654562  97.371   <2e-16 ***
DP           -0.009912   0.030489  -0.325    0.745
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.49 on 690 degrees of freedom
Multiple R-squared:  0.0001532,     Adjusted R-squared:  -0.001296
F-statistic: 0.1057 on 1 and 690 DF,  p-value: 0.7452
```

```
> summary(lm(HR ~ DP, subject$C))

Call:
lm(formula = HR ~ DP, data = subject$Alex)

Residuals:
    Min      1Q  Median      3Q     Max
-38.462  -6.442   0.482   6.579  33.619

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 154.726967   0.911897 169.676   <2e-16 ***
DP           -0.008039   0.017620  -0.456    0.648
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.56 on 1305 degrees of freedom
Multiple R-squared:  0.0001595,     Adjusted R-squared:  -0.0006067
F-statistic: 0.2082 on 1 and 1305 DF,  p-value: 0.6483
```

```
> summary(lm(HR ~ DP, subject$D))

Call:
lm(formula = HR ~ DP, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-41.897  -7.250   0.290   8.753  35.297

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 163.12899    1.79076   91.09  < 2e-16 ***
DP            0.16544    0.03182    5.20 2.61e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.01 on 712 degrees of freedom
Multiple R-squared:  0.03659, Adjusted R-squared:  0.03523
F-statistic: 27.04 on 1 and 712 DF,  p-value: 2.607e-07
```

```
> summary(lm(HR ~ DP, subject$E))

Call:
lm(formula = HR ~ DP, data = subject$Matt)

Residuals:
    Min      1Q  Median      3Q     Max
-35.923  -9.681  -1.611   8.668  44.396

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 152.74802    1.31725 115.960  < 2e-16 ***
DP           -0.12339    0.02222  -5.554 3.33e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.23 on 1416 degrees of freedom
Multiple R-squared:  0.02132, Adjusted R-squared:  0.02063
F-statistic: 30.85 on 1 and 1416 DF,  p-value: 3.328e-08
```

```
> summary(lm(HR ~ DP, subject$F))

Call:
lm(formula = HR ~ DP, data = subject$Joe)

Residuals:
    Min      1Q  Median      3Q     Max
-47.892  -8.034   0.213   8.847  37.728

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 163.9704     1.8627  88.029  < 2e-16 ***
DP           -0.1240     0.0309  -4.015 6.49e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.78 on 830 degrees of freedom
Multiple R-squared:  0.01905, Adjusted R-squared:  0.01787
F-statistic: 16.12 on 1 and 830 DF,  p-value: 6.489e-05
```

```
> summary(lm(HR ~ DP, subject$G))

Call:
lm(formula = HR ~ DP, data = subject$Adam)

Residuals:
    Min      1Q  Median      3Q     Max
-42.274 -12.286  -2.733  12.271  42.846

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 159.69917    1.71799   92.96  < 2e-16 ***
DP           -0.11064    0.02919   -3.79 0.000157 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.78 on 1419 degrees of freedom
Multiple R-squared:  0.01002, Adjusted R-squared:  0.009324
F-statistic: 14.36 on 1 and 1419 DF,  p-value: 0.0001569
```

```
> summary(lm(HR ~ DP, subject$H))

Call:
lm(formula = HR ~ DP, data = subject$Grant)

Residuals:
    Min      1Q  Median      3Q     Max
-39.019 -10.302  -1.849   8.605  30.073

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 155.21413    1.67526  92.651   <2e-16 ***
DP           -0.03414    0.03308  -1.032    0.302
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.39 on 867 degrees of freedom
Multiple R-squared:  0.001227,Adjusted R-squared:  7.475e-05
F-statistic: 1.065 on 1 and 867 DF,  p-value: 0.3024
```

For all but three of the data subjects the DP predictor presents a significant p-value well below the 0.01 threshold. The three failed data subjects are well above this limit and suggest no statistical relationship between HR and DP in their specific cases. As was mentioned in the expectations and limitations above this could likely be due to the increasingly small sample size of the data available for each data subject. The total numbers of observations range between 669-1421 depending on the subject, which is not a large enough number to filter out any large-scale variability that may be present from the data that was sampled.

The gender and individual subsets that do present significant p-values for the DP predictor variable, generate β_1 values between 0.09 to 0.16 which align with the values calculated for the dataset as a whole. This suggests that the model generated for the dataset as a whole would be acceptable for calculating the effect of DP on HR for each individual.

```
> summary(lm(Pace ~ DP, gender$F))

Call:
lm(formula = Pace ~ DP, data = gender$F)

Residuals:
     Min       1Q   Median       3Q      Max
-308.633  -26.379   -0.878   25.635  251.122

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 415.10939    2.57178  161.41   <2e-16 ***
DP            0.75047    0.04733   15.86   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 42.19 on 3380 degrees of freedom
Multiple R-squared:  0.06923, Adjusted R-squared:  0.06895
F-statistic: 251.4 on 1 and 3380 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ DP, gender$M))

Call:
lm(formula = Pace ~ DP, data = gender$M)

Residuals:
    Min      1Q  Median      3Q     Max
-128.11  -16.01    3.24   20.76  316.03

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 391.44494    2.25262 173.773  < 2e-16 ***
DP            0.24351    0.03898   6.247 4.58e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.93 on 4538 degrees of freedom
Multiple R-squared:  0.008525,Adjusted R-squared:  0.008307
F-statistic: 39.02 on 1 and 4538 DF,  p-value: 4.578e-10
```

```
> summary(lm(Pace ~ DP, subject$A))

Call:
lm(formula = Pace ~ DP, data = subject$Jordan)

Residuals:
     Min       1Q   Median       3Q      Max
-119.711  -14.626    0.578   17.070  129.289

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 441.57291    4.98078  88.655  < 2e-16 ***
DP            0.50851    0.08717   5.833 8.47e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.1 on 667 degrees of freedom
Multiple R-squared:  0.04854, Adjusted R-squared:  0.04711
F-statistic: 34.03 on 1 and 667 DF,  p-value: 8.469e-09
```

```
> summary(lm(Pace ~ DP, subject$B))

Call:
lm(formula = Pace ~ DP, data = subject$Molly)

Residuals:
     Min       1Q   Median       3Q      Max
-217.445  -18.386   -1.856   20.518  137.436

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 416.8844     5.5860  74.630   <2e-16 ***
DP            0.9405     0.1029   9.137   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38.8 on 690 degrees of freedom
Multiple R-squared:  0.1079,  Adjusted R-squared:  0.1066
F-statistic: 83.48 on 1 and 690 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ DP, subject$C))

Call:
lm(formula = Pace ~ DP, data = subject$Alex)

Residuals:
     Min       1Q   Median       3Q      Max
-123.447  -11.451   -0.164   13.832  191.286

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 410.59889    2.34285 175.256  < 2e-16 ***
DP            0.25192    0.04527   5.565 3.18e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.12 on 1305 degrees of freedom
Multiple R-squared:  0.02318, Adjusted R-squared:  0.02243
F-statistic: 30.97 on 1 and 1305 DF,  p-value: 3.179e-08
```

```
> summary(lm(Pace ~ DP, subject$D))

Call:
lm(formula = Pace ~ DP, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-338.51  -20.69    0.40   22.59  220.31

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 450.1553     6.4885  69.377  < 2e-16 ***
DP            0.6471     0.1153   5.613 2.85e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43.52 on 712 degrees of freedom
Multiple R-squared:  0.04238, Adjusted R-squared:  0.04103
F-statistic: 31.51 on 1 and 712 DF,  p-value: 2.848e-08
```

```
> summary(lm(Pace ~ DP, subject$E))

Call:
lm(formula = Pace ~ DP, data = subject$Matt)

Residuals:
     Min       1Q   Median       3Q      Max
-126.142  -16.853    2.316   19.665  186.238

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 413.57353    3.52916  117.19   <2e-16 ***
DP           -0.01727    0.05952   -0.29    0.772
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.43 on 1416 degrees of freedom
Multiple R-squared:  5.944e-05,     Adjusted R-squared:  -0.0006467
F-statistic: 0.08417 on 1 and 1416 DF,  p-value: 0.7718
```

```
> summary(lm(Pace ~ DP, subject$F))

Call:
lm(formula = Pace ~ DP, data = subject$Joe)

Residuals:
     Min       1Q   Median       3Q      Max
-121.623  -11.646    3.159   15.816  163.903

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 395.38034    4.97995  79.394  < 2e-16 ***
DP            0.30921    0.08261   3.743 0.000194 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34.16 on 830 degrees of freedom
Multiple R-squared:  0.0166,  Adjusted R-squared:  0.01541
F-statistic: 14.01 on 1 and 830 DF,  p-value: 0.0001944
```

```
> summary(lm(Pace ~ DP, subject$G))

Call:
lm(formula = Pace ~ DP, data = subject$Adam)

Residuals:
     Min       1Q   Median       3Q      Max
-117.911  -17.249    3.561   18.596  169.336

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 366.44659    3.77200   97.15  < 2e-16 ***
DP            0.52816    0.06409    8.24 3.86e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.84 on 1419 degrees of freedom
Multiple R-squared:  0.04567, Adjusted R-squared:  0.045
F-statistic:  67.9 on 1 and 1419 DF,  p-value: 3.86e-16
```

```
> summary(lm(Pace ~ DP, subject$H))

Call:
lm(formula = Pace ~ DP, data = subject$Grant)

Residuals:
    Min      1Q  Median      3Q     Max
-118.89  -19.54    2.22   25.01  316.43

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 402.55426    6.37818  63.114   <2e-16 ***
DP           -0.08532    0.12595  -0.677    0.498
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 54.79 on 867 degrees of freedom
Multiple R-squared:  0.000529,Adjusted R-squared:  -0.0006238
F-statistic: 0.4589 on 1 and 867 DF,  p-value: 0.4983
```

In the case of Pace as a response variable only two individuals didn't not show significance in DP as a predictor with p-values well above the upper threshold of 0.05.

Those with significant p-values for the DP predictor generated β_1 values between 0.24 and 0.94 with healthy displacement across this range. This doesn't fully align with the dataset as a whole at 0.27 but does offer even greater credence to the use of DP as a predictor of Pace.

Multiple Linear Regression

Using multiple predictors in a linear regression will offer the opportunity to understand the interaction effects that are occurring between the HR and Pace variables which seem to be very much intertwined from analysis of the above regression models.

Hypothesis:

There is a statistically significant synergistic interaction relationship between the response value DP and the multiplied variables HR and Pace.

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: \beta_1 = \beta_2 = \beta_3 \neq 0$

```
> summary(lm(DP ~ HR*Pace, dataset))

    Call:
    lm(formula = DP ~ HR * Pace, data = dataset)

    Residuals:
        Min      1Q  Median      3Q     Max
    -58.446 -12.227   1.682  13.662  26.398

    Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
    (Intercept) 45.0437174 15.5638627   2.894  0.00381 **
    HR          -0.0154863  0.0958264  -0.162  0.87162
    Pace         0.0384350  0.0362086   1.061  0.28850
    HR:Pace     -0.0000736  0.0002233  -0.330  0.74174
    ---
    Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Residual standard error: 15.5 on 7918 degrees of freedom
    Multiple R-squared:  0.009771,      Adjusted R-squared:  0.009396
    F-statistic: 26.04 on 3 and 7918 DF,  p-value: < 2.2e-16
```

It is immediately clear to see from this model that the p-values returned suggest little confidence in the hypothesis. In fact no synergistic relationship appears to exist between the two predictor variables (HR*Pace) on the response (DP).

It is interesting that individually the DP metric appears to have a predictive effect on the performance metrics, HR and Pace, as its value changes. But to see conversely, that there is no evidence of a synergistic relationship between HR and Pace, that could predict a change in the DP values is peculiar.

It suggests that it doesn’t actually make any real sense to use the performance metrics to predict the value of a weather value. Therefore a different approach was used below using the DP weather metric multiplied separately by one of the performance metrics at a time in order to predict the other. This should offer a much better insight of the interaction effect that exists between the two variables that both individually show a cause-effect relationship with the same predictor value.

```
> summary(lm(HR ~ DP*Pace, dataset))

Call:
lm(formula = HR ~ DP * Pace, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-71.431 -10.879  -0.205  10.248  56.330

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.865e+02  5.293e+00  35.234  < 2e-16 ***
DP          -3.198e-01  9.313e-02  -3.434 0.000599 ***
Pace        -6.766e-02  1.246e-02  -5.428 5.85e-08 ***
DP:Pace      6.522e-04  2.182e-04   2.989 0.002808 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.89 on 7918 degrees of freedom
Multiple R-squared:  0.01467, Adjusted R-squared:  0.0143
F-statistic:  39.3 on 3 and 7918 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ DP*HR, dataset))

Call:
lm(formula = Pace ~ DP * HR, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-273.009  -26.932   -4.902   27.969  283.491

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 532.289870  20.744700  25.659  < 2e-16 ***
DP           -0.990642   0.364996  -2.714 0.006660 **
HR           -0.775103   0.133089  -5.824 5.97e-09 ***
DP:HR         0.008073   0.002349   3.437 0.000591 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.28 on 7918 degrees of freedom
Multiple R-squared:  0.0198,  Adjusted R-squared:  0.01942
F-statistic:  53.3 on 3 and 7918 DF,  p-value: < 2.2e-16
```

These are definitely much more indicative models to use. There is evidence of synergies between the variables DP*Pace in respect to HR and DP*HR in respect to Pace, with p-values below the threshold of 0.1 in all cases, even when referring to the F-Statistic. However the r-squared values are suggesting just under 2% of the variability can be accounted for in the predictors values by variance in the response which hurts the statistical relevance suggested by the p-values.

The above method is applied to models of HR and Pace with respect to each other and HI in the hopes of establishing a synergistic cause-effect relationship to HI to support the effect from the DP variable.

```
> summary(lm(HR ~ HI*Pace, dataset))

Call:
lm(formula = HR ~ HI * Pace, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-70.988 -10.586  -0.182  10.066  57.647

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.856e+02  6.132e+00  30.261  < 2e-16 ***
HI          -2.570e-01  9.358e-02  -2.747 0.006034 **
Pace        -5.321e-02  1.430e-02  -3.722 0.000199 ***
HI:Pace      3.321e-04  2.179e-04   1.524 0.127491
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.8 on 7918 degrees of freedom
Multiple R-squared:  0.02655, Adjusted R-squared:  0.02618
F-statistic: 71.98 on 3 and 7918 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ HI*HR, dataset))

Call:
lm(formula = Pace ~ HI * HR, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-273.241  -26.925   -5.358   27.141  280.999

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 515.954643  24.157107  21.358  < 2e-16 ***
HI           -0.566641   0.362931  -1.561 0.118495
HR           -0.596381   0.153875  -3.876 0.000107 ***
HI:HR         0.003931   0.002325   1.691 0.090867 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.47 on 7918 degrees of freedom
Multiple R-squared:  0.01212, Adjusted R-squared:  0.01174
F-statistic: 32.37 on 3 and 7918 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ DP*HI, dataset))

Call:
lm(formula = HR ~ DP * HI, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-72.000 -10.322  -0.197   9.780  58.849

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 149.346422   2.231750  66.919  < 2e-16 ***
DP            0.433021   0.050784   8.527  < 2e-16 ***
HI           -0.006523   0.038884  -0.168    0.867
DP:HI        -0.004574   0.000722  -6.336 2.49e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.8 on 7918 degrees of freedom
Multiple R-squared:  0.02682, Adjusted R-squared:  0.02645
F-statistic: 72.75 on 3 and 7918 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ DP*HI, dataset))

Call:
lm(formula = Pace ~ DP * HI, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-283.877  -25.512   -2.985   26.591  293.169

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 423.334902   7.303454  57.964  < 2e-16 ***
DP            0.512257   0.166192   3.082 0.002061 **
HI           -0.468791   0.127249  -3.684 0.000231 ***
DP:HI         0.001425   0.002363   0.603 0.546523
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.44 on 7918 degrees of freedom
Multiple R-squared:  0.01351, Adjusted R-squared:  0.01314
F-statistic: 36.14 on 3 and 7918 DF,  p-value: < 2.2e-16
```

Support for these models is a little more sporadic than for the DP based models above. Each of these models returns at least one β value with a corresponding p-value well outside of the threshold of 0.05. Suggesting that the synergy effect may not exist when considering the HI in conjunction with the performance values or DP itself as a predictor.

--

Below are models for the gender and individual subsets using the performance metrics of HR and Pace as response variables with synergies calculated between DP and the unused performance metric in each case. These models follow the trend of the full dataset with only four individuals having one of the one of their predictors p-values grow outside the range of significance.

```
> summary(lm(HR~DP*Pace, gender$F))

Call:
lm(formula = HR ~ DP * Pace, data = gender$F)

Residuals:
    Min      1Q  Median      3Q     Max
-77.827  -7.960  -0.090   7.922  45.559

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.142e+02  8.442e+00  13.532  < 2e-16 ***
DP           9.656e-01  1.474e-01   6.551 6.60e-11 ***
Pace         9.071e-02  1.885e-02   4.812 1.56e-06 ***
DP:Pace     -1.897e-03  3.248e-04  -5.839 5.75e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.88 on 3378 degrees of freedom
Multiple R-squared:  0.02547, Adjusted R-squared:  0.02461
F-statistic: 29.43 on 3 and 3378 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, gender$M))

Call:
lm(formula = HR ~ DP * Pace, data = gender$M)

Residuals:
    Min      1Q  Median      3Q     Max
-49.809  -9.097  -1.205   8.397  58.184

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.964e+02  6.531e+00  30.068  < 2e-16 ***
DP           5.900e-01  1.203e-01   4.905 9.69e-07 ***
Pace        -1.017e-01  1.613e-02  -6.306 3.14e-10 ***
DP:Pace     -1.610e-03  2.957e-04  -5.443 5.51e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.06 on 4536 degrees of freedom
Multiple R-squared:  0.2653,  Adjusted R-squared:  0.2648
F-statistic: 545.9 on 3 and 4536 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$A))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Jordan)

Residuals:
    Min      1Q  Median      3Q     Max
-49.950  -4.668   0.911   6.085  30.709

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.304e+02  2.510e+01   5.195 2.73e-07 ***
DP           1.518e+00  4.136e-01   3.671 0.000261 ***
Pace         4.281e-02  5.371e-02   0.797 0.425723
DP:Pace     -2.850e-03  8.788e-04  -3.243 0.001243 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.45 on 665 degrees of freedom
Multiple R-squared:  0.1533,  Adjusted R-squared:  0.1495
F-statistic: 40.14 on 3 and 665 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$B))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Molly)

Residuals:
    Min      1Q  Median      3Q     Max
-79.506  -5.701   0.533   6.234  29.693

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.425e+02  1.725e+01   8.261 7.41e-16 ***
DP           1.393e+00  3.027e-01   4.602 4.97e-06 ***
Pace         2.561e-02  3.690e-02   0.694    0.488
DP:Pace     -2.715e-03  6.366e-04  -4.264 2.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.28 on 688 degrees of freedom
Multiple R-squared:  0.2014,  Adjusted R-squared:  0.1979
F-statistic: 57.83 on 3 and 688 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$C))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Alex)

Residuals:
    Min      1Q  Median      3Q     Max
-39.149  -5.127   0.136   4.942  34.106

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.769e+02  1.272e+01  13.907  < 2e-16 ***
DP           1.284e+00  2.215e-01   5.796 8.49e-09 ***
Pace        -5.849e-02  3.017e-02  -1.939   0.0527 .
DP:Pace     -2.921e-03  5.222e-04  -5.595 2.69e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.613 on 1303 degrees of freedom
Multiple R-squared:  0.3353,  Adjusted R-squared:  0.3338
F-statistic: 219.1 on 3 and 1303 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$D))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-43.614  -6.535   0.668   8.088  23.131

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.385e+02  1.852e+01   7.477 2.25e-13 ***
DP           1.513e+00  3.265e-01   4.634 4.27e-06 ***
Pace         4.324e-02  3.816e-02   1.133    0.258
DP:Pace     -2.625e-03  6.662e-04  -3.941 8.93e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.05 on 710 degrees of freedom
Multiple R-squared:  0.1869,  Adjusted R-squared:  0.1835
F-statistic: 54.42 on 3 and 710 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$E))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Matt)

Residuals:
    Min      1Q  Median      3Q     Max
-40.580  -7.215  -1.263   6.724  45.399

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.391e+02  1.260e+01  11.037  < 2e-16 ***
DP           1.507e+00  2.253e-01   6.690 3.20e-11 ***
Pace         3.280e-02  3.031e-02   1.082    0.279
DP:Pace     -3.947e-03  5.424e-04  -7.278 5.59e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.41 on 1414 degrees of freedom
Multiple R-squared:  0.2732,  Adjusted R-squared:  0.2717
F-statistic: 177.2 on 3 and 1414 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$F))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Joe)

Residuals:
    Min      1Q  Median      3Q     Max
-48.366  -8.051   0.137   8.421  37.965

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.438e+02  2.260e+01   6.362 3.29e-10 ***
DP           6.825e-01  3.638e-01   1.876   0.0610 .
Pace         4.643e-02  5.486e-02   0.846   0.3976
DP:Pace     -1.904e-03  8.791e-04  -2.166   0.0306 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.54 on 828 degrees of freedom
Multiple R-squared:  0.05789, Adjusted R-squared:  0.05448
F-statistic: 16.96 on 3 and 828 DF,  p-value: 1.068e-10
```

```
> summary(lm(HR~DP*Pace, subject$G))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Adam)

Residuals:
    Min      1Q  Median      3Q     Max
-52.796  -9.162  -2.242   8.130  67.441

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.966e+02  1.323e+01  14.863  < 2e-16 ***
DP           9.684e-01  2.424e-01   3.994 6.83e-05 ***
Pace        -1.070e-01  3.435e-02  -3.114  0.00188 **
DP:Pace     -2.463e-03  6.225e-04  -3.957 7.97e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.27 on 1417 degrees of freedom
Multiple R-squared:  0.2852,  Adjusted R-squared:  0.2837
F-statistic: 188.4 on 3 and 1417 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR~DP*Pace, subject$H))

Call:
lm(formula = HR ~ DP * Pace, data = subject$Grant)

Residuals:
    Min      1Q  Median      3Q     Max
-34.796  -6.920   1.135   6.697  43.088

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.060e+02  8.537e+00  24.130  < 2e-16 ***
DP           5.406e-01  1.691e-01   3.198  0.00144 **
Pace        -1.266e-01  2.091e-02  -6.055 2.09e-09 ***
DP:Pace     -1.462e-03  4.143e-04  -3.528  0.00044 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.435 on 865 degrees of freedom
Multiple R-squared:  0.5718,  Adjusted R-squared:  0.5703
F-statistic:   385 on 3 and 865 DF,  p-value: < 2.2e-16
```

--

```
> summary(lm(Pace ~ DP*HR, gender$F))

Call:
lm(formula = Pace ~ DP * HR, data = gender$F)

Residuals:
     Min       1Q   Median       3Q      Max
-308.033  -26.716   -1.092   25.978  248.623

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 252.422078  33.242609   7.593 4.01e-14 ***
DP            4.229182   0.593218   7.129 1.23e-12 ***
HR            1.017032   0.208361   4.881 1.10e-06 ***
DP:HR        -0.021679   0.003702  -5.857 5.18e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.94 on 3378 degrees of freedom
Multiple R-squared:  0.08078, Adjusted R-squared:  0.07996
F-statistic: 98.95 on 3 and 3378 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ DP*HR, gender$M))

Call:
lm(formula = Pace ~ DP * HR, data = gender$M)

Residuals:
     Min       1Q   Median       3Q      Max
-132.632  -17.896   -1.054   17.386  275.291

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 563.372846  18.761684  30.028  < 2e-16 ***
DP            0.877989   0.335526   2.617  0.00891 **
HR           -1.080880   0.122675  -8.811  < 2e-16 ***
DP:HR        -0.005152   0.002208  -2.333  0.01969 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.4 on 4536 degrees of freedom
Multiple R-squared:  0.2589,  Adjusted R-squared:  0.2584
F-statistic: 528.1 on 3 and 4536 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$A))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Jordan)

Residuals:
     Min       1Q   Median       3Q      Max
-155.731  -13.870    2.549   16.455   93.216

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 165.569160  61.092627   2.710   0.0069 **
DP            8.582917   1.109504   7.736 3.82e-14 ***
HR            1.694529   0.382752   4.427 1.12e-05 ***
DP:HR        -0.049770   0.006918  -7.194 1.70e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.11 on 665 degrees of freedom
Multiple R-squared:  0.2252,  Adjusted R-squared:  0.2217
F-statistic: 64.43 on 3 and 665 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$B))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Molly)

Residuals:
     Min       1Q   Median       3Q      Max
-215.973  -15.991    0.225   18.256  150.840

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 252.767700  69.025398   3.662  0.00027 ***
DP            8.327474   1.246950   6.678 4.99e-11 ***
HR            1.041470   0.431041   2.416  0.01594 *
DP:HR        -0.046383   0.007794  -5.951 4.24e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34.31 on 688 degrees of freedom
Multiple R-squared:  0.3045,  Adjusted R-squared:  0.3014
F-statistic: 100.4 on 3 and 688 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$C))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Alex)

Residuals:
     Min       1Q   Median       3Q      Max
-102.098  -12.450    0.927   11.379  145.084

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 361.628822  29.195309  12.387   <2e-16 ***
DP            5.418483   0.527141  10.279   <2e-16 ***
HR            0.315610   0.188218   1.677   0.0938 .
DP:HR        -0.033452   0.003397  -9.846   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.6 on 1303 degrees of freedom
Multiple R-squared:  0.3811,  Adjusted R-squared:  0.3797
F-statistic: 267.4 on 3 and 1303 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$D))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-333.19  -19.06   -0.16   20.71  224.99

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 269.606617  84.390303   3.195  0.00146 **
DP            8.327508   1.527924   5.450 6.95e-08 ***
HR            0.985243   0.491970   2.003  0.04559 *
DP:HR        -0.043297   0.008848  -4.893 1.23e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.8 on 710 degrees of freedom
Multiple R-squared:  0.2011,  Adjusted R-squared:  0.1977
F-statistic: 59.57 on 3 and 710 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$E))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Matt)

Residuals:
     Min       1Q   Median       3Q      Max
-119.031  -13.873   -1.637   12.809  181.734

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 393.858152  31.076829  12.674  < 2e-16 ***
DP            3.839748   0.550839   6.971 4.82e-12 ***
HR            0.178141   0.208870   0.853    0.394
DP:HR        -0.027323   0.003732  -7.322 4.08e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30.55 on 1414 degrees of freedom
Multiple R-squared:  0.2578,  Adjusted R-squared:  0.2562
F-statistic: 163.7 on 3 and 1414 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$F))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Joe)

Residuals:
     Min       1Q   Median       3Q      Max
-123.614  -12.812    2.582   16.132  154.771

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 292.28237   58.00999   5.038 5.76e-07 ***
DP            3.49649    0.98894   3.536 0.000429 ***
HR            0.69108    0.37084   1.864 0.062736 .
DP:HR        -0.02096    0.00636  -3.296 0.001022 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.39 on 828 degrees of freedom
Multiple R-squared:  0.06249, Adjusted R-squared:  0.05909
F-statistic:  18.4 on 3 and 828 DF,  p-value: 1.462e-11
```

```
> summary(lm(Pace~DP*HR, subject$G))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Adam)

Residuals:
     Min       1Q   Median       3Q      Max
-121.076  -16.833   -1.149   16.560  198.687

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 481.734533  25.334965  19.015  < 2e-16 ***
DP            1.673530   0.456585   3.665 0.000256 ***
HR           -0.704019   0.163421  -4.308 1.76e-05 ***
DP:HR        -0.008326   0.002968  -2.805 0.005095 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.41 on 1417 degrees of freedom
Multiple R-squared:  0.3071,  Adjusted R-squared:  0.3057
F-statistic: 209.4 on 3 and 1417 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace~DP*HR, subject$H))

Call:
lm(formula = Pace ~ DP * HR, data = subject$Grant)

Residuals:
     Min       1Q   Median       3Q      Max
-141.400  -24.309    3.785   19.211  251.019

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.516e+02  4.901e+01  17.377   <2e-16 ***
DP          -2.800e-01  9.413e-01  -0.297    0.766
HR          -2.894e+00  3.183e-01  -9.091   <2e-16 ***
DP:HR        6.333e-04  6.124e-03   0.103    0.918
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.18 on 865 degrees of freedom
Multiple R-squared:  0.5653,  Adjusted R-squared:  0.5638
F-statistic: 374.9 on 3 and 865 DF,  p-value: < 2.2e-16
```

Similar to earlier testing of Pace as a response variable, it has again shown that it is much more likely to return greater p-value significance from its predictor values than when HR is used as a response. There are only three predictors from two individuals of the 24 total calculated from the eight individuals sampled that have p-values which grew beyond the significance threshold. This shows great alignment with the dataset as a whole.

Predictive Analysis

Linear Regression

Due to the lack of strong correlations in the dataset the performance metrics will be tested as response variables against the weather metrics as predictors. Validation methods will be used to narrow down if any of these variables have a measurable effect on the performance metrics HR and Pace.

This hypothesis will be tested for the full dataset, each gender and each data subject individually. Each for two different response variables of HR and Pace.

Hypothesis:

There is a statistically significant relationship between the response variable and the selected predictor variables.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$$

$$H_1: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 \neq 0$$

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation, dataset))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation, dataset))
```

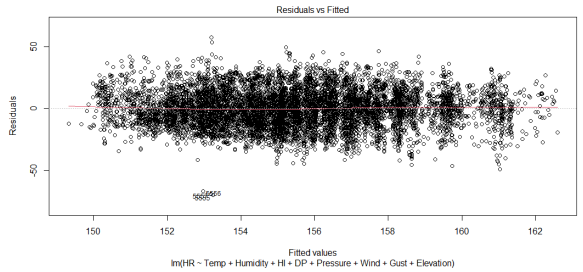
```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-70.239 -10.329  -0.215   9.757  57.798

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 127.100687  11.449354  11.101  < 2e-16 ***
Temp          1.439368   0.651788   2.208   0.0272 *
Humidity      0.299162   0.049162   6.085 1.22e-09 ***
HI           -1.076746   0.583506  -1.845   0.0650 .
DP           -0.388287   0.081807  -4.746 2.11e-06 ***
Pressure      0.004059   0.007994   0.508   0.6117
Wind         -0.008197   0.035592  -0.230   0.8179
Gust         -0.017644   0.025461  -0.693   0.4884
Elevation     0.003897   0.005487   0.710   0.4775
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.79 on 7913 degrees of freedom
Multiple R-squared:  0.02864, Adjusted R-squared:  0.02765
F-statistic: 29.16 on 8 and 7913 DF,  p-value: < 2.2e-16
```

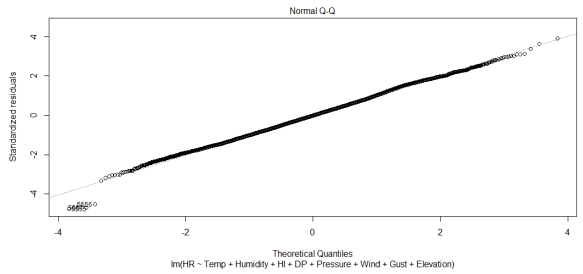
```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-287.993  -25.113   -2.646   26.362  294.293

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 195.06519   37.29020   5.231 1.73e-07 ***
Temp          7.53092    2.12285   3.548 0.000391 ***
Humidity      1.21322    0.16012   7.577 3.94e-14 ***
HI           -5.46930    1.90046  -2.878 0.004014 **
DP           -1.41621    0.26644  -5.315 1.09e-07 ***
Pressure      0.08370    0.02604   3.215 0.001311 **
Wind          0.07916    0.11592   0.683 0.494679
Gust         -0.07008    0.08293  -0.845 0.398094
Elevation     0.08489    0.01787   4.750 2.07e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48.18 on 7913 degrees of freedom
Multiple R-squared:  0.02468, Adjusted R-squared:  0.0237
F-statistic: 25.03 on 8 and 7913 DF,  p-value: < 2.2e-16
```

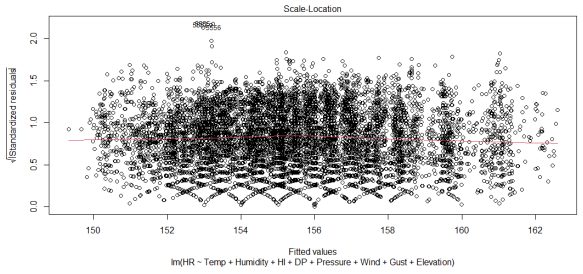
It is interesting to see from this initial regression model that both wind speed variables (Wind and Gust) are negligible on both sides when it comes to predicting a response effect on either HR or Pace.

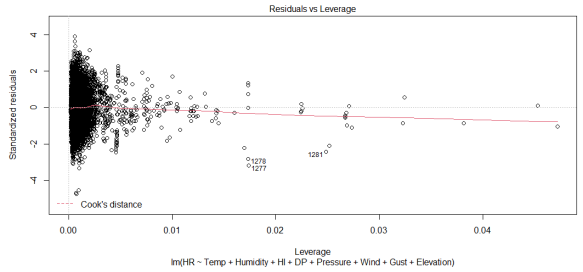
Outliers

```
> plot(hr.lm)
```

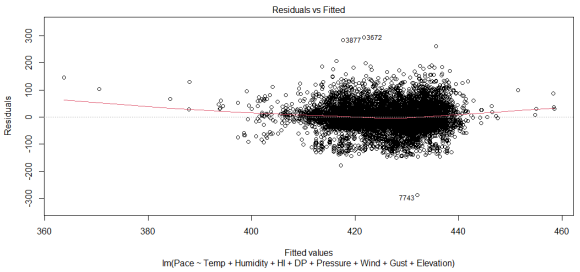


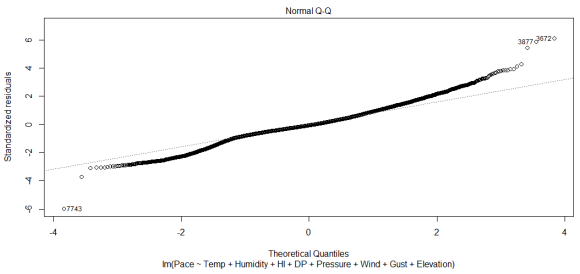


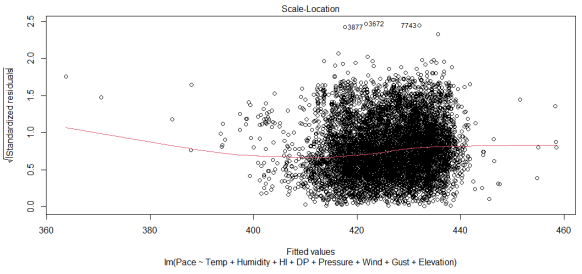


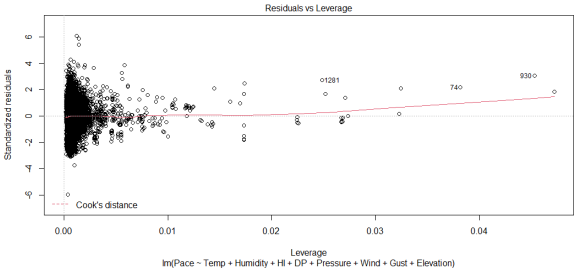



```
> plot(pace.lm)
```









The residual plots for the response variables HR and Pace are used to identify outliers in the dataset. These plots show linearity to a very acceptable level with little evidence of high leverage points. The Pace plots show some leverage but not to a degree that any of the data points should be considered an outlier and removed from the set. This suggests that there are not any outliers of concern in the dataset that would skew the data analysis. This outcome was expected due to the collection and cleaning methodologies used to build the dataset with conscious thought given to avoiding possible outliers.

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
      + Elevation , gender$F))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
      + Elevation , gender$F))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = gender$F)

Residuals:
    Min      1Q  Median      3Q     Max
-77.462  -7.575   0.018   7.790  48.454

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 94.8443566 15.1550643   6.258 4.38e-10 ***
Temp         2.4226141  0.8712291   2.781  0.00545 **
Humidity     0.2723552  0.0600221   4.538 5.89e-06 ***
HI          -1.9387366  0.7827088  -2.477  0.01330 *
DP          -0.2163802  0.0953145  -2.270  0.02326 *
Pressure     0.0240273  0.0117548   2.044  0.04103 *
Wind         0.1946124  0.0481444   4.042 5.41e-05 ***
Gust         0.0006399  0.0392300   0.016  0.98699
Elevation   -0.0124611  0.0089278  -1.396  0.16288
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.86 on 3373 degrees of freedom
Multiple R-squared:  0.03001, Adjusted R-squared:  0.02771
F-statistic: 13.04 on 8 and 3373 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = gender$F)

Residuals:
     Min       1Q   Median       3Q      Max
-314.472  -24.021   -1.292   23.960  249.286

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 159.48867   49.09331   3.249 0.001171 **
Temp          2.20834    2.82226   0.782 0.433991
Humidity     -0.12174    0.19444  -0.626 0.531279
HI           -2.49123    2.53551  -0.983 0.325906
DP            1.17794    0.30876   3.815 0.000139 ***
Pressure      0.25660    0.03808   6.739 1.87e-11 ***
Wind          0.04650    0.15596   0.298 0.765615
Gust          0.16419    0.12708   1.292 0.196439
Elevation     0.06339    0.02892   2.192 0.028449 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 41.67 on 3373 degrees of freedom
Multiple R-squared:  0.09399, Adjusted R-squared:  0.09184
F-statistic: 43.74 on 8 and 3373 DF,  p-value: < 2.2e-16
```

Curiously the regression models, when applied solely to the female population of the dataset, introduce significance for variables that were showing p-values far above any threshold when viewed in the context of the whole dataset.

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , gender$M))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , gender$M))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = gender$M)

Residuals:
    Min      1Q  Median      3Q     Max
-47.814 -10.668  -1.164   9.962  45.495

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 167.067175  16.143518  10.349   <2e-16 ***
Temp          0.752195   0.887004   0.848    0.396
Humidity      0.045350   0.075376   0.602    0.547
HI           -0.859594   0.789544  -1.089    0.276
DP            0.029485   0.130068   0.227    0.821
Pressure     -0.012394   0.010426  -1.189    0.235
Wind         -0.092805   0.047742  -1.944    0.052 .
Gust          0.024987   0.031997   0.781    0.435
Elevation     0.008159   0.006611   1.234    0.217
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.06 on 4531 degrees of freedom
Multiple R-squared:  0.02525, Adjusted R-squared:  0.02353
F-statistic: 14.67 on 8 and 4531 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = gender$M)

Residuals:
    Min      1Q  Median      3Q     Max
-134.58  -17.34    2.03   21.38  318.69

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 276.404725  43.417632   6.366 2.13e-10 ***
Temp         10.456093   2.385577   4.383 1.20e-05 ***
Humidity      0.460481   0.202723   2.271  0.02316 *
HI           -8.990384   2.123462  -4.234 2.34e-05 ***
DP           -0.159804   0.349815  -0.457  0.64782
Pressure      0.001055   0.028041   0.038  0.96998
Wind          0.361876   0.128402   2.818  0.00485 **
Gust         -0.034629   0.086054  -0.402  0.68740
Elevation     0.078208   0.017779   4.399 1.11e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.49 on 4531 degrees of freedom
Multiple R-squared:  0.03115, Adjusted R-squared:  0.02944
F-statistic: 18.21 on 8 and 4531 DF,  p-value: < 2.2e-16
```

It seems bizarre that the full dataset showed no p-value significance for the inclusion of either of the wind speed variables in predicting response from either HR or Pace but that when predicting at the gender level Wind is generating p-values of significant levels in three out of four cases. In fact the male regression is not showing significance below a 0.05 threshold for any of the variables at predicting the response in HR which is startling.

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$A))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$A))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Jordan)

Residuals:
    Min      1Q  Median      3Q     Max
-41.489  -6.243   0.781   6.807  29.549

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 115.84036   58.94544   1.965   0.0498 *
Temp          8.14476    1.84040   4.426 1.13e-05 ***
Humidity      0.18653    0.18441   1.011   0.3122
HI           -7.60186    1.60845  -4.726 2.80e-06 ***
DP            0.41434    0.38433   1.078   0.2814
Pressure     -0.02912    0.04985  -0.584   0.5593
Wind         -0.07814    0.11345  -0.689   0.4912
Gust          0.16557    0.06944   2.384   0.0174 *
Elevation     0.00615    0.02169   0.284   0.7769
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.97 on 660 degrees of freedom
Multiple R-squared:  0.07339, Adjusted R-squared:  0.06215
F-statistic: 6.534 on 8 and 660 DF,  p-value: 3.231e-08
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Jordan)

Residuals:
     Min       1Q   Median       3Q      Max
-117.192  -14.702    1.301   15.111  130.638

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 750.94046  163.69486   4.587 5.37e-06 ***
Temp         -7.18445    5.11091  -1.406   0.1603
Humidity     -0.01211    0.51211  -0.024   0.9811
HI            8.11770    4.46677   1.817   0.0696 .
DP           -1.18110    1.06731  -1.107   0.2689
Pressure     -0.26460    0.13845  -1.911   0.0564 .
Wind         -0.31571    0.31505  -1.002   0.3167
Gust         -0.12154    0.19285  -0.630   0.5288
Elevation     0.06883    0.06023   1.143   0.2536
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.47 on 660 degrees of freedom
Multiple R-squared:  0.09675, Adjusted R-squared:  0.0858
F-statistic: 8.837 on 8 and 660 DF,  p-value: 1.571e-11
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$B))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$B))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Molly)

Residuals:
    Min      1Q  Median      3Q     Max
-73.278  -6.327   0.897   7.321  27.960

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 87.25487   46.04598   1.895  0.05852 .
Temp        -4.36651    1.81134  -2.411  0.01619 *
Humidity    -0.13739    0.23517  -0.584  0.55926
HI           3.80711    1.52552   2.496  0.01281 *
DP           0.11811    0.56582   0.209  0.83471
Pressure     0.11190    0.03713   3.013  0.00268 **
Wind         0.11230    0.10845   1.035  0.30080
Gust        -0.01665    0.07956  -0.209  0.83428
Elevation   -0.01052    0.01191  -0.883  0.37730
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.4 on 683 degrees of freedom
Multiple R-squared:  0.02652,   Adjusted R-squared:  0.01512
F-statistic: 2.326 on 8 and 683 DF,  p-value: 0.01815
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Molly)

Residuals:
     Min       1Q   Median       3Q      Max
-221.668  -17.399   -0.619   18.732  134.583

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 142.67732  154.37682   0.924   0.3557
Temp          6.76313    6.07283   1.114   0.2658
Humidity     -0.23229    0.78845  -0.295   0.7684
HI           -6.93671    5.11457  -1.356   0.1755
DP            1.83712    1.89701   0.968   0.3332
Pressure      0.25076    0.12450   2.014   0.0444 *
Wind         -0.20260    0.36361  -0.557   0.5776
Gust          0.58060    0.26674   2.177   0.0298 *
Elevation     0.09243    0.03992   2.315   0.0209 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 38.21 on 683 degrees of freedom
Multiple R-squared:  0.1435,    Adjusted R-squared:  0.1335
F-statistic:  14.3 on 8 and 683 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$C))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$C))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Alex)

Residuals:
    Min      1Q  Median      3Q     Max
-37.150  -6.696   0.286   6.441  32.935

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 139.70344   16.23562   8.605  < 2e-16 ***
Temp          1.93560    1.10700   1.749 0.080614 .
Humidity      0.23137    0.06743   3.431 0.000620 ***
HI           -1.43810    0.99896  -1.440 0.150221
DP           -0.32963    0.08878  -3.713 0.000213 ***
Pressure     -0.01769    0.01247  -1.419 0.156050
Wind          0.10271    0.07063   1.454 0.146119
Gust         -0.18696    0.06065  -3.083 0.002094 **
Elevation    -0.01343    0.01338  -1.003 0.315859
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.42 on 1298 degrees of freedom
Multiple R-squared:  0.03009,   Adjusted R-squared:  0.02411
F-statistic: 5.033 on 8 and 1298 DF,  p-value: 3.573e-06
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Alex)

Residuals:
     Min       1Q   Median       3Q      Max
-121.811  -11.300    0.148   11.863  194.270

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 489.21947   41.72945  11.724  < 2e-16 ***
Temp          0.50067    2.84527   0.176  0.86035
Humidity     -0.02251    0.17332  -0.130  0.89668
HI           -0.47658    2.56756  -0.186  0.85277
DP            0.35958    0.22817   1.576  0.11529
Pressure     -0.08662    0.03204  -2.704  0.00695 **
Wind          0.18847    0.18153   1.038  0.29933
Gust          0.33598    0.15588   2.155  0.03132 *
Elevation     0.15164    0.03440   4.408 1.13e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26.79 on 1298 degrees of freedom
Multiple R-squared:  0.05166,   Adjusted R-squared:  0.04581
F-statistic: 8.838 on 8 and 1298 DF,  p-value: 7.768e-12
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$D))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$D))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-39.252  -7.095   0.262   7.014  34.714

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 361.179128  53.360434   6.769 2.74e-11 ***
Temp          4.168973   1.820348   2.290  0.02230 *
Humidity     -0.051763   0.224587  -0.230  0.81778
HI           -4.431486   1.561145  -2.839  0.00466 **
DP            0.924657   0.512130   1.806  0.07142 .
Pressure     -0.219875   0.041352  -5.317 1.42e-07 ***
Wind          0.275148   0.090040   3.056  0.00233 **
Gust         -0.044875   0.069995  -0.641  0.52166
Elevation     0.004919   0.021684   0.227  0.82061
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.51 on 705 degrees of freedom
Multiple R-squared:  0.1239,    Adjusted R-squared:  0.114
F-statistic: 12.46 on 8 and 705 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-332.63  -20.56    1.13   22.02  214.91

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 821.29665  196.13022   4.188 3.18e-05 ***
Temp          0.31155    6.69082   0.047   0.9629
Humidity      0.18753    0.82549   0.227   0.8203
HI            1.38745    5.73810   0.242   0.8090
DP           -1.32055    1.88237  -0.702   0.4832
Pressure     -0.36456    0.15199  -2.399   0.0167 *
Wind         -1.71206    0.33095  -5.173 3.00e-07 ***
Gust         -0.03424    0.25727  -0.133   0.8941
Elevation     0.04188    0.07970   0.525   0.5994
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42.3 on 705 degrees of freedom
Multiple R-squared:  0.1039,    Adjusted R-squared:  0.0937
F-statistic: 10.21 on 8 and 705 DF,  p-value: 1.457e-13
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$E))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$E))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Matt)

Residuals:
    Min      1Q  Median      3Q     Max
-35.992  -9.605  -1.588   8.905  43.900

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 110.084089  26.698198   4.123 3.95e-05 ***
Temp          0.389139   1.411462   0.276   0.7828
Humidity      0.148088   0.121953   1.214   0.2248
HI           -0.137412   1.249033  -0.110   0.9124
DP           -0.388520   0.212246  -1.831   0.0674 .
Pressure      0.029450   0.015907   1.851   0.0643 .
Wind          0.059228   0.090231   0.656   0.5117
Gust         -0.028938   0.051436  -0.563   0.5738
Elevation     0.005008   0.007391   0.678   0.4982
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.22 on 1409 degrees of freedom
Multiple R-squared:  0.02693,   Adjusted R-squared:  0.0214
F-statistic: 4.873 on 8 and 1409 DF,  p-value: 5.983e-06
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Matt)

Residuals:
     Min       1Q   Median       3Q      Max
-120.227  -16.589    2.635   21.095  150.713

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 815.64534   68.75837  11.862  < 2e-16 ***
Temp         -3.83714    3.63507  -1.056 0.291337
Humidity     -0.73861    0.31408  -2.352 0.018826 *
HI            2.41752    3.21675   0.752 0.452452
DP            1.31837    0.54662   2.412 0.015998 *
Pressure     -0.32612    0.04097  -7.961 3.49e-15 ***
Wind         -0.17447    0.23238  -0.751 0.452890
Gust          0.36249    0.13247   2.736 0.006289 **
Elevation     0.07091    0.01904   3.725 0.000203 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 34.05 on 1409 degrees of freedom
Multiple R-squared:  0.08133,   Adjusted R-squared:  0.07611
F-statistic: 15.59 on 8 and 1409 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$F))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$F))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Joe)

Residuals:
    Min      1Q  Median      3Q     Max
-48.264  -7.358  -0.265   7.952  38.770

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.743e+02  6.746e+01   4.067 5.23e-05 ***
Temp         4.054e+00  1.689e+00   2.401  0.01659 *
Humidity     5.194e-02  1.900e-01   0.273  0.78461
HI          -4.384e+00  1.467e+00  -2.988  0.00289 **
DP           5.214e-01  3.767e-01   1.384  0.16668
Pressure    -1.282e-01  6.206e-02  -2.065  0.03920 *
Wind         2.111e-02  1.168e-01   0.181  0.85663
Gust        -2.875e-02  8.019e-02  -0.359  0.72000
Elevation   -5.375e-04  2.164e-02  -0.025  0.98019
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.47 on 823 degrees of freedom
Multiple R-squared:  0.07354,   Adjusted R-squared:  0.06454
F-statistic: 8.166 on 8 and 823 DF,  p-value: 1.148e-10
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Joe)

Residuals:
     Min       1Q   Median       3Q      Max
-121.599  -13.575    2.068   17.310  165.664

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 418.93895  181.42887   2.309   0.0212 *
Temp         10.20999    4.54211   2.248   0.0248 *
Humidity     -0.02452    0.51085  -0.048   0.9617
HI           -9.29114    3.94522  -2.355   0.0188 *
DP            0.61480    1.01303   0.607   0.5441
Pressure     -0.09625    0.16689  -0.577   0.5643
Wind         -0.38139    0.31414  -1.214   0.2251
Gust         -0.04080    0.21564  -0.189   0.8500
Elevation     0.12195    0.05819   2.096   0.0364 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.53 on 823 degrees of freedom
Multiple R-squared:  0.06027,   Adjusted R-squared:  0.05113
F-statistic: 6.598 on 8 and 823 DF,  p-value: 2.254e-08
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$G))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$G))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Adam)

Residuals:
    Min      1Q  Median      3Q     Max
-46.211 -12.258  -2.289  12.780  42.249

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 135.18160   28.06491   4.817 1.62e-06 ***
Temp         -1.08892    1.77099  -0.615   0.5387
Humidity      0.06983    0.13762   0.507   0.6119
HI            0.98949    1.60159   0.618   0.5368
DP           -0.18340    0.22757  -0.806   0.4204
Pressure      0.03011    0.01824   1.651   0.0989 .
Wind          0.15740    0.07511   2.096   0.0363 *
Gust          0.08679    0.06131   1.416   0.1571
Elevation     0.03488    0.01555   2.243   0.0250 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.6 on 1412 degrees of freedom
Multiple R-squared:  0.03534,   Adjusted R-squared:  0.02987
F-statistic: 6.466 on 8 and 1412 DF,  p-value: 2.732e-08
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Adam)

Residuals:
     Min       1Q   Median       3Q      Max
-128.571  -16.882    0.272   17.898  163.192

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 221.24081   59.58134   3.713 0.000213 ***
Temp         13.83591    3.75978   3.680 0.000242 ***
Humidity     -0.10357    0.29216  -0.355 0.723013
HI          -12.67060    3.40014  -3.726 0.000202 ***
DP            0.94074    0.48313   1.947 0.051713 .
Pressure      0.04259    0.03871   1.100 0.271426
Wind          0.03387    0.15946   0.212 0.831804
Gust         -0.13200    0.13016  -1.014 0.310699
Elevation     0.08200    0.03301   2.484 0.013104 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.25 on 1412 degrees of freedom
Multiple R-squared:  0.1306,    Adjusted R-squared:  0.1256
F-statistic:  26.5 on 8 and 1412 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$H))
> summary(lm(Pace ~ Temp + Humidity + HI + DP + Pressure + Wind + Gust
    + Elevation , subject$H))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Grant)

Residuals:
    Min      1Q  Median      3Q     Max
-38.893  -8.560  -1.109   7.787  34.280

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 518.467945  44.811710  11.570  < 2e-16 ***
Temp         -0.803991   1.922813  -0.418    0.676
Humidity     -0.104958   0.188999  -0.555    0.579
HI            0.422958   1.693517   0.250    0.803
DP            0.189661   0.338496   0.560    0.575
Pressure     -0.336611   0.034907  -9.643  < 2e-16 ***
Wind         -0.586521   0.118956  -4.931 9.84e-07 ***
Gust          0.092369   0.062493   1.478    0.140
Elevation    -0.008576   0.018346  -0.467    0.640
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.06 on 860 degrees of freedom
Multiple R-squared:  0.184,     Adjusted R-squared:  0.1764
F-statistic: 24.24 on 8 and 860 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + HI + DP + Pressure + Wind +
    Gust + Elevation, data = subject$Grant)

Residuals:
    Min      1Q  Median      3Q     Max
-129.33  -25.76    1.79   22.67  319.84

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 172.52084  184.81224   0.933  0.35083
Temp         14.56336    7.93006   1.836  0.06663 .
Humidity     -0.41533    0.77947  -0.533  0.59428
HI          -14.93581    6.98440  -2.138  0.03276 *
DP            2.11517    1.39602   1.515  0.13010
Pressure      0.14441    0.14396   1.003  0.31609
Wind          1.50292    0.49060   3.063  0.00226 **
Gust         -0.32283    0.25774  -1.253  0.21071
Elevation     0.04409    0.07566   0.583  0.56022
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 53.87 on 860 degrees of freedom
Multiple R-squared:  0.04186,   Adjusted R-squared:  0.03295
F-statistic: 4.696 on 8 and 860 DF,  p-value: 1.2e-05
```

The individual subsets are again showing great variability in the results returned as has been in the case in a number of the previous tests. Suggestion that their calculations are inaccurate on account of their small sample size and high variability.

Best Subset Selection

```
> library(leaps)
> hr.predictors <- dataset[c(7:12, 15:17)]
> pace.predictors <- dataset[c(6, 8:12, 15:17)]

> hrfit <- summary(regsubsets(HR ~ ., hr.predictors))
> pacefit <- summary(regsubsets(Pace ~ ., pace.predictors))
```

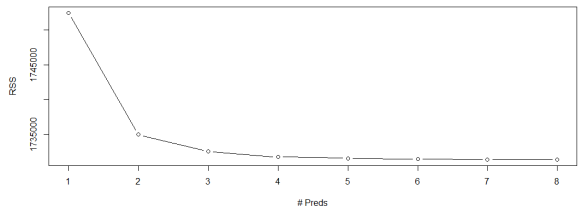
```
Subset selection object
Call: regsubsets.formula(HR ~ ., hr.predictors)
8 Variables  (and intercept)
           Forced in Forced out
Elevation      FALSE      FALSE
Temp           FALSE      FALSE
Humidity       FALSE      FALSE
DP             FALSE      FALSE
HI             FALSE      FALSE
Wind           FALSE      FALSE
Gust           FALSE      FALSE
Pressure       FALSE      FALSE
1 subsets of each size up to 8
Selection Algorithm: exhaustive
         Elevation Temp Humidity DP  HI  Wind Gust Pressure
1  ( 1 ) " "       "*"  " "      " " " " " "  " "  " "
2  ( 1 ) " "       " "  "*"      "*" " " " "  " "  " "
3  ( 1 ) " "       "*"  "*"      "*" " " " "  " "  " "
4  ( 1 ) " "       "*"  "*"      "*" "*" " "  " "  " "
5  ( 1 ) " "       "*"  "*"      "*" "*" " "  "*"  " "
6  ( 1 ) "*"       "*"  "*"      "*" "*" " "  "*"  " "
7  ( 1 ) "*"       "*"  "*"      "*" "*" " "  "*"  "*"
8  ( 1 ) "*"       "*"  "*"      "*" "*" "*"  "*"  "*"
```

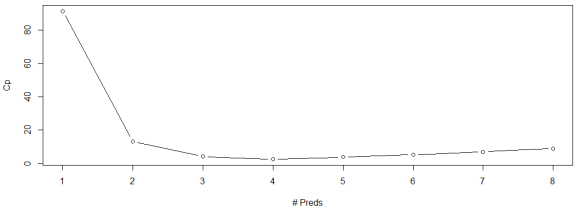
```
Subset selection object
Call: regsubsets.formula(Pace ~ ., pace.predictors)
8 Variables  (and intercept)
           Forced in Forced out
Elevation      FALSE      FALSE
Temp           FALSE      FALSE
Humidity       FALSE      FALSE
DP             FALSE      FALSE
HI             FALSE      FALSE
Wind           FALSE      FALSE
Gust           FALSE      FALSE
Pressure       FALSE      FALSE
1 subsets of each size up to 8
Selection Algorithm: exhaustive
         Elevation Temp Humidity DP  HI  Wind Gust Pressure
1  ( 1 ) " "       " "  "*"      " " " " " "  " "  " "
2  ( 1 ) "*"       " "  "*"      " " " " " "  " "  " "
3  ( 1 ) "*"       "*"  "*"      " " " " " "  " "  " "
4  ( 1 ) "*"       "*"  "*"      "*" " " " "  " "  " "
5  ( 1 ) "*"       "*"  "*"      "*" " " " "  " "  "*"
6  ( 1 ) "*"       "*"  "*"      "*" "*" " "  " "  "*"
7  ( 1 ) "*"       "*"  "*"      "*" "*" " "  "*"  "*"
8  ( 1 ) "*"       "*"  "*"      "*" "*" "*"  "*"  "*"
```

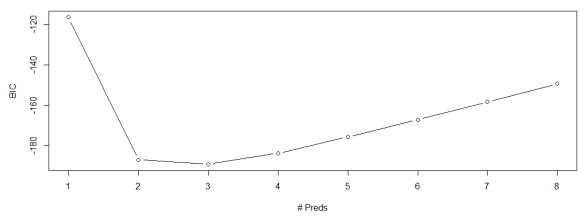
```
> summary(hrfit)$rsq
[1] 0.01681530 0.02663376 0.02800147 0.02843213
[5] 0.02853806 0.02859995 0.02862909 0.02863560
```

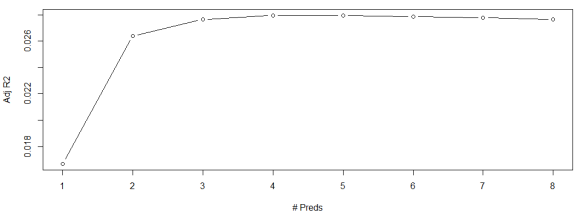
```
> summary(pacefit)$rsq
[1] 0.01525548 0.01811295 0.02000000 0.02207458
[5] 0.02354494 0.02458351 0.02462589 0.02468338
```

```
> plot(summary(hrfit)$rss, xlab="# Preds", ylab="RSS", type="b")
> plot(summary(hrfit)$cp, xlab="# Preds", ylab="Cp", type="b")
> plot(summary(hrfit)$bic, xlab="# Preds", ylab="BIC", type="b")
> plot(summary(hrfit)$adjr2, xlab="# Preds", ylab="Adj R2", type="b")
```

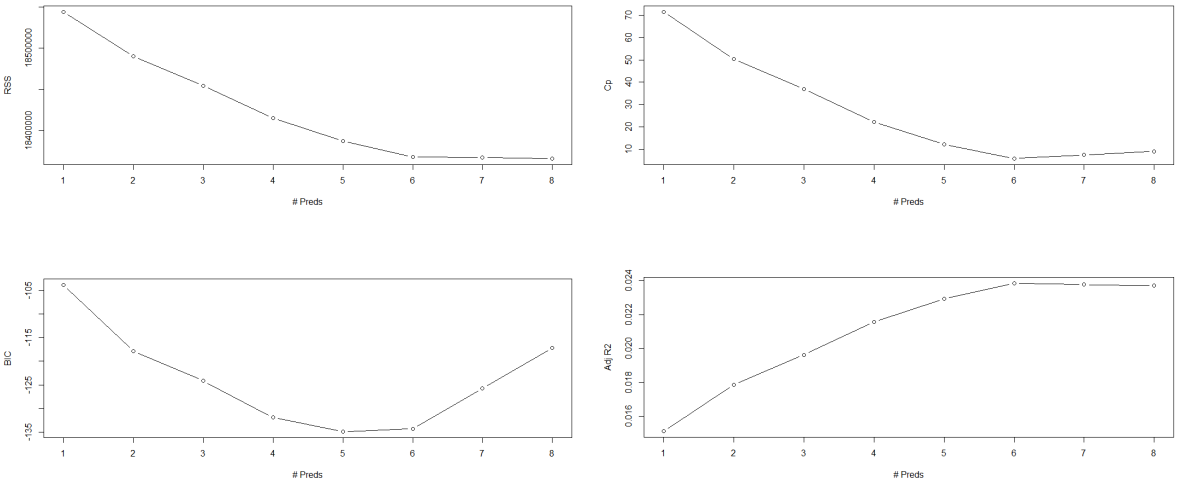








```
> plot(summary(pacefit)$rss, xlab="# Preds", ylab="RSS", type="b")
> plot(summary(pacefit)$cp, xlab="# Preds", ylab="Cp", type="b")
> plot(summary(pacefit)$bic, xlab="# Preds", ylab="BIC", type="b")
> plot(summary(pacefit)$adjr2, xlab="# Preds", ylab="Adj R2", type="b")
```




```
> which.min(summary(hrfit)$rss)
  [1] 8

> which.min(summary(hrfit)$cp)
  [1] 4

> which.min(summary(hrfit)$bic)
  [1] 3

> which.max(summary(hrfit)$adjr2)
  [1] 4
```

```
> which.min(summary(pacefit)$rss)
  [1] 8

> which.min(summary(pacefit)$cp)
  [1] 6

> which.min(summary(pacefit)$bic)
  [1] 5

> which.max(summary(pacefit)$adjr2)
  [1] 6
```

For (HR ~ .) it appears that the subset containing four variable components is the best subset, with two of the four error tests indicating so, the other two offered two separate subsets as the best option.
The variables within the chosen dataset are, Temp, Humidity, DP and HI. This is very pleasant to observe as this selection of variables is the exact weather metrics that were suspected from the beginning to have the greatest effect on the performance metrics.

For (Pace ~ .) it appears that the subset containing six variable components is the best subset, with two of the four error tests indicating so, the other two offered two separate subsets as the best option.
The variables within the chosen dataset are, Elevation, Temp, Humidity, DP, HI and Pressure. This is an interesting addition to the variable predictors found to best predict HR, the Elevation variable was added to the dataset in the hopes of understanding if it was measurable to calculate the effect of running up or down hill on Pace. The Pressure variable was included for the same reason, it is interesting to see that both of these variables are considered significant to the response for Pace but not HR.

Both best subset selections match my own visual inspection from the above linear regressions and so will be used for predictive analysis in all cases, the full dataset and the gender and individual subsets.

Hypothesis:

There is a statistically significant relationship between the response variable and the selected predictor variables.

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_1: \beta_1 = \beta_2 = \beta_3 = \beta_4 \neq 0$

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, hr.predictors))

Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = hr.predictors)

Residuals:
    Min      1Q  Median      3Q     Max
-70.162 -10.365  -0.250   9.838  57.821

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 130.94831    7.12313  18.384  < 2e-16 ***
Temp          1.44951    0.65149   2.225   0.0261 *
Humidity      0.29960    0.04720   6.347 2.32e-10 ***
DP           -0.38159    0.07317  -5.215 1.89e-07 ***
HI           -1.09114    0.58247  -1.873   0.0611 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.79 on 7917 degrees of freedom
Multiple R-squared:  0.02843,Adjusted R-squared:  0.02794
F-statistic: 57.92 on 4 and 7917 DF,  p-value: < 2.2e-16
```

The greatest significance in this case is from the predictors Humidity and DP both of which produce p-values well below the 0.01 threshold. While Temp is within the 0.05 threshold, HI is just above this but at only 0.06 could still be considered usable. The p-value for the F-statistic supports significance across all of the predictors being well below the 0.01 threshold at 2.2^{16}. The r-squared values are again only suggesting that about 2.8% of the variability in this model can be attributed to the variability in the response which dampens the statistical relevance of the formulas produced.

```
HR = 130.94831 + 1.44951(Temp) + 0.29960(Humidity)
               - 0.38159(DP)   - 1.09114(HI)
```

```
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure
    + Elevation, pace.predictors))

Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = pace.predictors)

Residuals:
     Min       1Q   Median       3Q      Max
-287.930  -25.067   -2.635   26.392  293.726

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 193.40425   37.21670   5.197 2.08e-07 ***
Temp          7.57605    2.12207   3.570 0.000359 ***
Humidity      1.21204    0.15840   7.652 2.22e-14 ***
DP           -1.41330    0.26355  -5.363 8.44e-08 ***
HI           -5.51418    1.89948  -2.903 0.003706 **
Pressure      0.08570    0.02590   3.309 0.000942 ***
Elevation     0.08487    0.01787   4.750 2.07e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.17 on 7915 degrees of freedom
Multiple R-squared:  0.02458,    Adjusted R-squared:  0.02384
F-statistic: 33.25 on 6 and 7915 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, gender$F))
> summary(lm(HR ~ Temp + Humidity + DP + HI, gender$M))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = gender$F)

Residuals:
    Min      1Q  Median      3Q     Max
-76.632  -7.739  -0.015   8.108  49.340

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 122.82417    9.31716  13.183  < 2e-16 ***
Temp          2.71817    0.86623   3.138 0.001716 **
Humidity      0.22706    0.05905   3.845 0.000123 ***
DP           -0.06995    0.08841  -0.791 0.428854
HI           -2.34581    0.77576  -3.024 0.002514 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.91 on 3377 degrees of freedom
Multiple R-squared:  0.02203,   Adjusted R-squared:  0.02087
F-statistic: 19.02 on 4 and 3377 DF,  p-value: 1.757e-15
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = gender$M)

Residuals:
    Min      1Q  Median      3Q     Max
-44.532 -10.753  -1.087  10.050  45.106

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 148.30295    9.91091  14.964   <2e-16 ***
Temp          0.90955    0.88493   1.028    0.304
Humidity      0.10410    0.06994   1.488    0.137
DP           -0.08957    0.11243  -0.797    0.426
HI           -0.89614    0.78901  -1.136    0.256
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.06 on 4535 degrees of freedom
Multiple R-squared:  0.02367,   Adjusted R-squared:  0.02281
F-statistic: 27.48 on 4 and 4535 DF,  p-value: < 2.2e-16
```

```
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    gender$F))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    gender$M))
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = gender$F)

Residuals:
     Min       1Q   Median       3Q      Max
-314.833  -24.167   -1.198   24.241  248.892

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 160.95974   49.05951   3.281  0.00105 **
Temp          1.91467    2.80036   0.684  0.49420
Humidity     -0.17197    0.19179  -0.897  0.36998
DP            1.24727    0.30427   4.099 4.24e-05 ***
HI           -2.29512    2.51440  -0.913  0.36142
Pressure      0.26201    0.03790   6.914 5.62e-12 ***
Elevation     0.06271    0.02892   2.168  0.03020 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.68 on 3375 degrees of freedom
Multiple R-squared:  0.09332,   Adjusted R-squared:  0.09171
F-statistic:  57.9 on 6 and 3375 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = gender$M)

Residuals:
     Min       1Q   Median       3Q      Max
-133.757  -17.018    1.951   21.365  315.059

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 281.54537   43.34663   6.495 9.18e-11 ***
Temp         10.13431    2.38442   4.250 2.18e-05 ***
Humidity      0.36749    0.20063   1.832   0.0671 .
DP           -0.03746    0.34768  -0.108   0.9142
HI           -8.81396    2.12334  -4.151 3.37e-05 ***
Pressure      0.00829    0.02782   0.298   0.7657
Elevation     0.07868    0.01779   4.422 1.00e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.53 on 4533 degrees of freedom
Multiple R-squared:  0.0291,    Adjusted R-squared:  0.02781
F-statistic: 22.64 on 6 and 4533 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$A))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$A))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Jordan)

Residuals:
    Min      1Q  Median      3Q     Max
-41.428  -6.280   0.762   6.944  29.894

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  97.4331    22.5218   4.326 1.75e-05 ***
Temp          7.2273     1.7787   4.063 5.42e-05 ***
Humidity      0.1099     0.1770   0.621    0.535
DP            0.5142     0.3728   1.379    0.168
HI           -6.8544     1.5534  -4.413 1.19e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.99 on 664 degrees of freedom
Multiple R-squared:  0.06471,   Adjusted R-squared:  0.05907
F-statistic: 11.48 on 4 and 664 DF,  p-value: 5.085e-09
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Jordan)

Residuals:
     Min       1Q   Median       3Q      Max
-117.963  -14.528    1.082   15.469  129.920

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 655.12826  148.96621   4.398 1.27e-05 ***
Temp         -6.83514    4.95351  -1.380    0.168
Humidity      0.18552    0.49407   0.375    0.707
DP           -1.55414    1.03914  -1.496    0.135
HI            8.17098    4.31981   1.892    0.059 .
Pressure     -0.19292    0.12596  -1.532    0.126
Elevation     0.06568    0.06017   1.092    0.275
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.47 on 662 degrees of freedom
Multiple R-squared:  0.09357,   Adjusted R-squared:  0.08535
F-statistic: 11.39 on 6 and 662 DF,  p-value: 3.877e-12
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$B))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$B))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Molly)

Residuals:
    Min      1Q  Median      3Q     Max
-75.107  -6.738   0.994   7.301  28.292

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 199.9093    27.1744   7.357  5.4e-13 ***
Temp         -3.6977     1.8008  -2.053   0.0404 *
Humidity     -0.1956     0.2352  -0.832   0.4059
DP            0.2551     0.5643   0.452   0.6514
HI            3.1035     1.5100   2.055   0.0402 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.45 on 687 degrees of freedom
Multiple R-squared:  0.01214,   Adjusted R-squared:  0.006392
F-statistic: 2.111 on 4 and 687 DF,  p-value: 0.07779
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Molly)

Residuals:
    Min      1Q  Median      3Q     Max
-225.32  -17.51   -0.69   18.21  142.67

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 144.15195  150.29117   0.959   0.3378
Temp          5.80733    6.06038   0.958   0.3383
Humidity     -0.30604    0.78923  -0.388   0.6983
DP            1.85825    1.89749   0.979   0.3278
HI           -6.08381    5.10579  -1.192   0.2339
Pressure      0.25980    0.11992   2.167   0.0306 *
Elevation     0.09259    0.03999   2.316   0.0209 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 38.29 on 685 degrees of freedom
Multiple R-squared:  0.1375,    Adjusted R-squared:   0.13
F-statistic:  18.2 on 6 and 685 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$C))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$C))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Alex)

Residuals:
    Min      1Q  Median      3Q     Max
-37.019  -6.918   0.280   6.486  34.207

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 121.51248   11.39474  10.664  < 2e-16 ***
Temp          2.01360    1.10314   1.825 0.068180 .
Humidity      0.23318    0.06734   3.463 0.000552 ***
DP           -0.35649    0.08778  -4.061 5.17e-05 ***
HI           -1.48182    0.99417  -1.491 0.136331
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.46 on 1302 degrees of freedom
Multiple R-squared:  0.02066,   Adjusted R-squared:  0.01765
F-statistic: 6.866 on 4 and 1302 DF,  p-value: 1.812e-05
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Alex)

Residuals:
     Min       1Q   Median       3Q      Max
-122.238  -11.839   -0.248   12.685  193.130

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 476.64544   40.96888  11.634  < 2e-16 ***
Temp          0.62382    2.84094   0.220   0.8262
Humidity     -0.04239    0.17362  -0.244   0.8071
DP            0.40152    0.22822   1.759   0.0788 .
HI           -0.63727    2.56279  -0.249   0.8037
Pressure     -0.07086    0.03101  -2.285   0.0225 *
Elevation     0.15086    0.03448   4.375 1.31e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26.86 on 1300 degrees of freedom
Multiple R-squared:  0.04527,   Adjusted R-squared:  0.04087
F-statistic: 10.27 on 6 and 1300 DF,  p-value: 3.871e-11
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$D))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$D))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-41.086  -7.437   0.315   8.547  36.662

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 116.1461    25.6437   4.529 6.94e-06 ***
Temp          4.3151     1.8607   2.319   0.0207 *
Humidity      0.2735     0.2148   1.273   0.2034
DP            0.0653     0.4931   0.132   0.8947
HI           -3.8415     1.5997  -2.401   0.0166 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.96 on 709 degrees of freedom
Multiple R-squared:  0.04847,   Adjusted R-squared:  0.04311
F-statistic:  9.03 on 4 and 709 DF,  p-value: 4.072e-07
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Sally)

Residuals:
    Min      1Q  Median      3Q     Max
-334.92  -19.99   -0.27   23.02  218.25

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 396.82121  186.67271   2.126   0.0339 *
Temp          0.55457    6.73300   0.082   0.9344
Humidity      1.04953    0.80702   1.301   0.1939
DP           -3.08145    1.86766  -1.650   0.0994 .
HI            2.87012    5.81040   0.494   0.6215
Pressure     -0.03552    0.14369  -0.247   0.8048
Elevation     0.05522    0.08145   0.678   0.4980
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.28 on 707 degrees of freedom
Multiple R-squared:  0.05936,   Adjusted R-squared:  0.05138
F-statistic: 7.436 on 6 and 707 DF,  p-value: 9.769e-08
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$E))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$E))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Matt)

Residuals:
    Min      1Q  Median      3Q     Max
-37.011  -9.711  -1.606   8.896  44.011

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 151.87951   15.31330   9.918   <2e-16 ***
Temp          0.01401    1.38648   0.010    0.992
Humidity      0.02933    0.10635   0.276    0.783
DP           -0.12321    0.16507  -0.746    0.456
HI           -0.03227    1.23704  -0.026    0.979
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.22 on 1413 degrees of freedom
Multiple R-squared:  0.02354,   Adjusted R-squared:  0.02077
F-statistic: 8.515 on 4 and 1413 DF,  p-value: 8.661e-07
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Matt)

Residuals:
     Min       1Q   Median       3Q      Max
-120.917  -16.360    3.154   20.938  155.377

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 824.40669   68.71392  11.998  < 2e-16 ***
Temp         -3.52399    3.62559  -0.972 0.331229
Humidity     -0.78294    0.31378  -2.495 0.012704 *
DP            1.42570    0.54486   2.617 0.008975 **
HI            2.02992    3.20098   0.634 0.526082
Pressure     -0.33324    0.04048  -8.231 4.17e-16 ***
Elevation     0.07144    0.01908   3.744 0.000188 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 34.13 on 1411 degrees of freedom
Multiple R-squared:  0.07554,   Adjusted R-squared:  0.07161
F-statistic: 19.22 on 6 and 1411 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$F))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$F))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Joe)

Residuals:
    Min      1Q  Median      3Q     Max
-47.748  -7.205  -0.293   7.916  39.429

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 148.42546   21.38041   6.942 7.81e-12 ***
Temp          3.55755    1.65061   2.155  0.03143 *
Humidity      0.01898    0.17623   0.108  0.91424
DP            0.56904    0.35749   1.592  0.11182
HI           -3.95399    1.44548  -2.735  0.00636 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.47 on 827 degrees of freedom
Multiple R-squared:  0.06826,   Adjusted R-squared:  0.06375
F-statistic: 15.15 on 4 and 827 DF,  p-value: 5.884e-12
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Joe)

Residuals:
     Min       1Q   Median       3Q      Max
-122.853  -13.952    2.337   17.067  164.834

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 330.82452  169.88736   1.947   0.0518 .
Temp         11.19183    4.47551   2.501   0.0126 *
Humidity      0.23391    0.47513   0.492   0.6226
DP            0.17746    0.96291   0.184   0.8538
HI           -9.77898    3.91882  -2.495   0.0128 *
Pressure     -0.03930    0.16178  -0.243   0.8081
Elevation     0.12163    0.05818   2.090   0.0369 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.53 on 825 degrees of freedom
Multiple R-squared:  0.05806,   Adjusted R-squared:  0.05121
F-statistic: 8.475 on 6 and 825 DF,  p-value: 6.018e-09
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$G))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$G))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Adam)

Residuals:
    Min      1Q  Median      3Q     Max
-45.302 -12.116  -2.447  12.571  42.329

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 176.46182   18.90724   9.333   <2e-16 ***
Temp         -1.40511    1.76962  -0.794    0.427
Humidity     -0.03848    0.13129  -0.293    0.769
DP            0.01216    0.19948   0.061    0.951
HI            1.11022    1.59690   0.695    0.487
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.7 on 1416 degrees of freedom
Multiple R-squared:  0.02177,   Adjusted R-squared:  0.01901
F-statistic: 7.878 on 4 and 1416 DF,  p-value: 2.8e-06
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Adam)

Residuals:
     Min       1Q   Median       3Q      Max
-127.857  -17.278    0.035   18.087  163.802

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 218.49382   59.45983   3.675 0.000247 ***
Temp         13.71518    3.73524   3.672 0.000250 ***
Humidity     -0.09991    0.28984  -0.345 0.730374
DP            0.94057    0.48265   1.949 0.051521 .
HI          -12.55984    3.37936  -3.717 0.000210 ***
Pressure      0.04570    0.03840   1.190 0.234219
Elevation     0.08143    0.03300   2.468 0.013712 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.24 on 1414 degrees of freedom
Multiple R-squared:  0.1299,    Adjusted R-squared:  0.1262
F-statistic: 35.18 on 6 and 1414 DF,  p-value: < 2.2e-16
```

```
> summary(lm(HR ~ Temp + Humidity + DP + HI, subject$H))
> summary(lm(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    subject$H))
```

```
Call:
lm(formula = HR ~ Temp + Humidity + DP + HI, data = subject$Grant)

Residuals:
    Min      1Q  Median      3Q     Max
-40.022 -10.228  -1.453   8.158  32.844

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 115.8593    25.0508   4.625 4.32e-06 ***
Temp          2.1040     2.0558   1.023   0.3064
Humidity      0.4489     0.1952   2.299   0.0217 *
DP           -0.5855     0.3506  -1.670   0.0953 .
HI           -1.5343     1.8152  -0.845   0.3982
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.1 on 864 degrees of freedom
Multiple R-squared:  0.04456,   Adjusted R-squared:  0.04013
F-statistic: 10.07 on 4 and 864 DF,  p-value: 5.684e-08
```

```
Call:
lm(formula = Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = subject$Grant)

Residuals:
     Min       1Q   Median       3Q      Max
-125.663  -23.807    1.802   24.590  304.629

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 142.55490  185.19261   0.770   0.4416
Temp         13.74249    7.95562   1.727   0.0845 .
Humidity     -0.97722    0.75910  -1.287   0.1983
DP            3.20272    1.35090   2.371   0.0180 *
HI          -15.11170    7.01454  -2.154   0.0315 *
Pressure      0.23098    0.14201   1.626   0.1042
Elevation     0.04597    0.07601   0.605   0.5455
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 54.12 on 862 degrees of freedom
Multiple R-squared:  0.03059,   Adjusted R-squared:  0.02384
F-statistic: 4.534 on 6 and 862 DF,  p-value: 0.0001549
```

The gender and individual subsets have responded in a similar way to the earlier regression testing, in that the predictor variables do not align significantly with those considered significant to the dataset as a whole.

Due to the lack of alignment in significance between the best subset predictors of the full dataset and the gender/individual subsets, only the full dataset response will be cross validated.

Validation Set Approach

```
> library(caret)

> hr.part <- createDataPartition(y=hr.predictors$HR, p=0.7, list=FALSE)

> hr.training <- hr.predictors[hr.part, ]
> hr.test <- hr.predictors[-hr.part, ]

> hr.lm.fit <- lm(HR ~ Temp + Humidity + DP + HI, hr.training)

> summary(hr.lm.fit)$r.squared
    [1] 0.03025492
```

This value indicates that a little over 3% of the variance from the predictor values in the training model can be accounted for by the variance in the response. Obviously this level of confidence is unacceptable and offers little support for the hypothesis that HR can be confidently determined by the variability in the selected weather metrics. Despite the indication from the models produced that such a relationship does exist.

```
> mean((hr.test$HR - predict(hr.lm.fit, hr.test))^2)
    [1] 224.7184

> sqrt(224.7184)
    [1] 14.99061
--
> hr.part2 <- createDataPartition(y = hr.predictors$HR, p = 0.7,
    list = FALSE)

> hr.training2 <- hr.predictors[hr.part2, ]
> hr.test2 <- hr.predictors[-hr.part2, ]

> hr.lm.fit2 <- lm(HR ~ Temp + Humidity + DP + HI, hr.training2)

> summary(hr.lm.fit2)$r.squared
    [1] 0.0323924

> mean((hr.test2$HR - predict(hr.lm.fit2, hr.test2))^2)
    [1] 226.732

> sqrt(226.732)
    [1] 15.05762
```

These error margins from these validation methods are difficult to see. The mean squared error value is incomprehensible as there isn't even 225 bpm variability in a person's HR from minimum resting to maximum exertion. The more neutral rooted mean squared error supposes close to 15 bpm in variability which is a much more acceptable amount of variability considering this dataset includes both male and female subjects of differing collegiate abilities.

```
> pace.part <- createDataPartition(y = pace.predictors$Pace, p = 0.7,
    list=FALSE)

> pace.training <- pace.predictors[pace.part, ]
> pace.test <- pace.predictors[-pace.part, ]

> pace.lm.fit <- lm(Pace ~ Temp + Humidity + DP + HI + Pressure
    + Elevation, pace.training)

> summary(pace.lm.fit)$r.squared
    [1] 0.02356728

> mean((pace.test$Pace - predict(pace.lm.fit, pace.test))^2)
    [1] 2315.831

> sqrt(2315.831)
    [1] 48.12308
```

```
> pace.part2 <- createDataPartition(y = pace.predictors$Pace, p = 0.7,
    list = FALSE)
> pace.training2 <- pace.predictors[pace.part2, ]
> pace.test2 <- pace.predictors[-pace.part2, ]
> pace.lm.fit2 <- lm(Pace ~ Temp + Humidity + DP + HI + Pressure
    + Elevation, pace.training2)
> summary(pace.lm.fit2)$r.squared
    [1] 0.02547088
> mean((pace.test2$Pace - predict(pace.lm.fit2, pace.test2))^2)
    [1] 2246.389
> sqrt(2246.389)
    [1] 47.39609
```

The validation set approach used here for the response variable Pace, offers even lower r-squared values at approximately 2.5%, than did the HR response validation. This suggests that a staggering 97.5% of the variance in the predictor values is simply unaccounted for by variance in the response.

Leave-One-Out Cross-Validation

```
> loocv <- trainControl(method = "LOOCV")
> train(HR ~ Temp + Humidity + DP + HI,
    data = hr.predictors, method = "lm", trControl = loocv)
    Linear Regression

    7922 samples
       4 predictor

    No pre-processing
    Resampling: Leave-One-Out Cross-Validation
    Summary of sample sizes: 7921, 7921, 7921, 7921, 7921, 7921, ...
    Resampling results:

      RMSE      Rsquared    MAE
      14.79364  0.02732911  11.85061

    Tuning parameter 'intercept' was held constant at a value of TRUE

> train(Pace ~ Temp + Humidity + DP + HI + Pressure + Elevation,
    data = pace.predictors, method = "lm", trControl = loocv)
    Linear Regression

    7922 samples
       6 predictor

    No pre-processing
    Resampling: Leave-One-Out Cross-Validation
    Summary of sample sizes: 7921, 7921, 7921, 7921, 7921, 7921, ...
    Resampling results:

      RMSE      Rsquared   MAE
      48.19829  0.0227571  35.67709

    Tuning parameter 'intercept' was held constant at a value of TRUE
```

Root mean squared error (RMSE) measures average differences between the predictions made by the model and the actual observations. The lower the RMSE, the more closely a model can predict the actual observations.

R-squared is a measure of correlation between predictions made by the model and the actual observations. The higher the R-squared, the more closely a model can predict actual observations.

Mean absolute error (MAE) is the average absolute difference between the predictions made by the model and the actual observations. The lower the MAE, the more closely a model can predict the actual observations.

The values presented here are similar to those estimated in the validation set approach above. RMSE is high at 14.8 and 48.2 points respectively for the response variables HR and Pace. and the accuracy probabilities of 0.027 and 0.023 from R-squared are startlingly below the common threshold of 0.7. These error calculations all point to a lack of accuracy in the model at predicting the performance metrics based on the predictor variables.

Ridge Regression

```
  > library(glmnet)
  > grid <- 10 ^ seq(4, -2, length = 100)
--
  > hr.train.mtrx <- model.matrix(HR ~ ., data = hr.training)
  > hr.test.mtrx <- model.matrix(HR ~ ., data = hr.test)

  > hr.cv.ridge <- cv.glmnet(hr.train.mtrx, hr.training$HR, alpha = 0,
      lambda = grid, thresh = 1e-12)

  > hr.ridge <- hr.cv.ridge$lambda.min
      [1] 0.01

  > hr.ridge.fit <- glmnet(hr.train.mtrx, hr.training$HR, alpha = 0,
      lambda = grid, thresh = 1e-12)

  > hr.predict.ridge <- predict(hr.ridge.fit, s = hr.ridge,
      newx = hr.test.mtrx)

  > predict(hr.ridge.fit, s = hr.ridge, type = "coefficients")

        10 x 1 sparse Matrix of class "dgCMatrix"
                              s1
        (Intercept) 133.391126270
        (Intercept)   .
        Elevation     0.002848369
        Temp          0.185216521
        Humidity      0.248331249
        DP           -0.406381910
        HI            0.073769838
        Wind         -0.019813446
        Gust         -0.012481478
        Pressure      0.009661854
```

Formula

```
HR = 143.7658293 + 0.3262518(Temp)       + 0.2224023(Humidity)
                 - 0.3168839(DP)         - 0.1263059(HI)
                 - 0.019813446(Wind)     - 0.012481478(Gust)
                 + 0.009661854(Pressure) + 0.002848369(Elevation)
```

```
> mean((hr.test$HR - hr.predict.ridge)^2)
    [1] 220.3672

> sqrt(220.3672)
    [1] 14.84477
```

--

```
> pace.train.mtrx <- model.matrix(Pace ~ ., data = pace.training)
> pace.test.mtrx <- model.matrix(Pace ~ ., data = pace.test)

> pace.cv.ridge <- cv.glmnet(pace.train.mtrx, pace.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> pace.ridge <- pace.cv.ridge$lambda.min
    [1] 0.01

> pace.ridge.fit <- glmnet(pace.train.mtrx, pace.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> pace.predict.ridge <- predict(pace.ridge.fit, s = pace.ridge,
    newx = pace.test.mtrx)

> predict(pace.ridge.fit, s = pace.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 243.84596702
    (Intercept)   .
    Elevation     0.07861091
    Temp          4.09811768
    Humidity      0.88885883
    DP           -1.02002327
    HI           -2.70464319
    Wind          0.01619371
    Gust         -0.03388748
    Pressure      0.08176007
```

Formula

```
HR = 243.84596702 + 4.09811768(Temp)     + 0.88885883(Humidity)
                  - 1.02002327(DP)       - 2.70464319(HI)
                  + 0.01619371(Wind)     - 0.03388748(Gust)
                  + 0.08176007(Pressure) + 0.07861091(Elevation)
```

```
> mean((pace.test$Pace - pace.predict.ridge)^2)
    [1] 2314.398

> sqrt(2314.398)
    [1] 48.10819
```

The MSE and RMSE values produced here are very similar to the values cross validated using the linear model and best subset selection. The ridge makes use of all of the weather variables as predictors which is likely unnecessary due to the minimal effect of certain variables in this prediction.

--

The functions used to generate the training and testing sets for the gender and subject subsets are provided in the appendix.

```
> hrW.cv.ridge <- cv.glmnet(hrW.train.mtrx, hrW.training$HR, alpha = 0,
lambda = grid, thresh = 1e-12)

> hrW.ridge <- hrW.cv.ridge$lambda.min
    [1] 0.01

> hrW.ridge.fit <- glmnet(hrW.train.mtrx, hrW.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrW.predict.ridge <- predict(hrW.ridge.fit, s = hrW.ridge,
    newx = hrW.test.mtrx)

> predict(hrW.ridge.fit, s = hrW.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 111.744852521
    (Intercept)   .
    Elevation    -0.013327208
    Temp          0.742988130
    Humidity      0.163850777
    DP           -0.146522442
    HI           -0.478114557
    Wind          0.202427437
    Gust          0.004827955
    Pressure      0.025965030

> mean((hrW.test$HR - hrW.predict.ridge)^2)
    [1] 167.6622

> sqrt(167.6622)
    [1] 12.94844
```

--

```
> hrM.cv.ridge <- cv.glmnet(hrM.train.mtrx, hrM.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrM.ridge <- hrM.cv.ridge$lambda.min
    [1] 2.31013

> hrM.ridge.fit <- glmnet(hrM.train.mtrx, hrM.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrM.predict.ridge <- predict(hrM.ridge.fit, s = hrM.ridge,
    newx = hrM.test.mtrx)
```

```
> predict(hrM.ridge.fit, s = hrM.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 164.596233000
    (Intercept)   .
    Elevation     0.006962509
    Temp         -0.067684411
    Humidity      0.036827295
    DP           -0.033675269
    HI           -0.058519561
    Wind         -0.028413722
    Gust          0.013504863
    Pressure     -0.004836235

> mean((hrM.test$HR - hrM.predict.ridge)^2)
    [1] 227.3133

> sqrt(227.3133)
    [1] 15.07691
--
> paceW.cv.ridge <- cv.glmnet(paceW.train.mtrx, paceW.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceW.ridge <- paceW.cv.ridge$lambda.min
    [1] 0.1629751

> paceW.ridge.fit <- glmnet(paceW.train.mtrx, paceW.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceW.predict.ridge <- predict(paceW.ridge.fit, s = paceW.ridge,
    newx = paceW.test.mtrx)

> predict(paceW.ridge.fit, s = paceW.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 146.89948905
    (Intercept)   .
    Elevation     0.05993247
    Temp         -0.07040759
    Humidity     -0.05066561
    DP            0.74294015
    HI           -0.04149279
    Wind          0.12119710
    Gust          0.04901368
    Pressure      0.27627316

> mean((paceW.test$Pace - paceW.predict.ridge)^2)
    [1] 1921.65

> sqrt(1921.65)
    [1] 43.83663
```

```
> paceM.cv.ridge <- cv.glmnet(paceM.train.mtrx, paceM.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceM.ridge <- paceM.cv.ridge$lambda.min
    [1] 0.01

> paceM.ridge.fit <- glmnet(paceM.train.mtrx, paceM.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceM.predict.ridge <- predict(paceM.ridge.fit, s = paceM.ridge,
    newx = paceM.test.mtrx)

> predict(paceM.ridge.fit, s = paceM.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 343.081949307
    (Intercept)   .
    Elevation     0.093751789
    Temp          5.552451867
    Humidity      0.052516081
    DP            0.231345052
    HI           -4.809237371
    Wind          0.149405775
    Gust         -0.014775589
    Pressure     -0.007322454

> mean((paceM.test$Pace - paceM.predict.ridge)^2)
    [1] 1582.451

> sqrt(1582.451)
    [1] 39.78003
--
> hrA.cv.ridge <- cv.glmnet(hrA.train.mtrx, hrA.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrA.ridge <- hrA.cv.ridge$lambda.min
    [1] 14.17474

> hrA.ridge.fit <- glmnet(hrA.train.mtrx, hrA.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrA.predict.ridge <- predict(hrA.ridge.fit, s = hrA.ridge,
    newx = hrA.test.mtrx)

> predict(hrA.ridge.fit, s = hrA.ridge, type = "coefficients")

      10 x 1 sparse Matrix of class "dgCMatrix"
                             s1
      (Intercept) 183.574443339
      (Intercept)   .
      Elevation     0.005453988
      Temp          0.031399194
      Humidity     -0.006452383
      DP            0.020878865
      HI            0.026195354
      Wind          0.014450285
      Gust          0.078714107
      Pressure     -0.027553388
```

```
> mean((hrA.test$HR - hrA.predict.ridge)^2)
    [1] 132.6129

> sqrt(132.6129)
    [1] 11.51577
```

--

```
> hrB.cv.ridge <- cv.glmnet(hrB.train.mtrx, hrB.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrB.ridge <- hrB.cv.ridge$lambda.min
    [1] 10000

> hrB.ridge.fit <- glmnet(hrB.train.mtrx, hrB.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrB.predict.ridge <- predict(hrB.ridge.fit, s = hrB.ridge,
    newx = hrB.test.mtrx)

> predict(hrB.ridge.fit, s = hrB.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept)  1.605829e+02
    (Intercept)  .
    Elevation   -1.103269e-05
    Temp        -2.985627e-05
    Humidity     5.233988e-05
    DP          -8.115369e-06
    HI          -2.430898e-05
    Wind         4.951221e-05
    Gust        -6.428298e-05
    Pressure     1.815149e-05

> mean((hrB.test$HR - hrB.predict.ridge)^2)
    [1] 137.6286

> sqrt(137.6286)
    [1] 11.73152
```

--

```
> hrC.cv.ridge <- cv.glmnet(hrC.train.mtrx, hrC.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrC.ridge <- hrC.cv.ridge$lambda.min
    [1] 0.02656088

> hrC.ridge.fit <- glmnet(hrC.train.mtrx, hrC.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrC.predict.ridge <- predict(hrC.ridge.fit, s = hrC.ridge,
    newx = hrC.test.mtrx)
```

```
> predict(hrC.ridge.fit, s = hrC.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 165.01215803
    (Intercept)   .
    Elevation    -0.01434266
    Temp          0.21177515
    Humidity      0.13965380
    DP           -0.28484177
    HI            0.06285002
    Wind          0.19329041
    Gust         -0.22965842
    Pressure     -0.02428025

> mean((hrC.test$HR - hrC.predict.ridge)^2)
    [1] 103.8443

> sqrt(103.8443)
    [1] 10.1904
--
> hrD.cv.ridge <- cv.glmnet(hrD.train.mtrx, hrD.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrD.ridge <- hrD.cv.ridge$lambda.min
    [1] 0.01

> hrD.ridge.fit <- glmnet(hrD.train.mtrx, hrD.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrD.predict.ridge <- predict(hrD.ridge.fit, s = hrD.ridge,
    newx = hrD.test.mtrx)

> predict(hrD.ridge.fit, s = hrD.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 390.956539660
    (Intercept)   .
    Elevation    -0.002163391
    Temp          0.519013568
    Humidity     -0.181533813
    DP            0.735134096
    HI           -0.972644187
    Wind          0.277537306
    Gust         -0.080937420
    Pressure     -0.215963619

> mean((hrD.test$HR - hrD.predict.ridge)^2)
    [1] 148.5612

> sqrt(148.5612)
    [1] 12.18857
--
> hrE.cv.ridge <- cv.glmnet(hrE.train.mtrx, hrE.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrE.ridge <- hrE.cv.ridge$lambda.min
    [1] 8.111308
```

```
> hrE.ridge.fit <- glmnet(hrE.train.mtrx, hrE.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrE.predict.ridge <- predict(hrE.ridge.fit, s = hrE.ridge,
    newx = hrE.test.mtrx)

> predict(hrE.ridge.fit, s = hrE.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 143.616752465
    (Intercept)   .
    Elevation     0.007202814
    Temp         -0.036234240
    Humidity     -0.010730651
    DP           -0.040782333
    HI           -0.033614461
    Wind          0.061547295
    Gust         -0.034671705
    Pressure      0.009232491

> mean((hrE.test$HR - hrE.predict.ridge)^2)
    [1] 180.9057

> sqrt(180.9057)
    [1] 13.45012
```

--

```
> hrF.cv.ridge <- cv.glmnet(hrF.train.mtrx, hrF.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrF.ridge <- hrF.cv.ridge$lambda.min
    [1] 0.01

> hrF.ridge.fit <- glmnet(hrF.train.mtrx, hrF.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrF.predict.ridge <- predict(hrF.ridge.fit, s = hrF.ridge,
    newx = hrF.test.mtrx)

> predict(hrF.ridge.fit, s = hrF.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 309.736947669
    (Intercept)   .
    Elevation    -0.005831289
    Temp          1.136938236
    Humidity     -0.096129465
    DP            0.534262229
    HI           -1.742821675
    Wind         -0.043168301
    Gust         -0.025005676
    Pressure     -0.134030647

> mean((hrF.test$HR - hrF.predict.ridge)^2)
    [1] 148.9587

> sqrt(148.9587)
    [1] 12.20486
```

--

```
> hrG.cv.ridge <- cv.glmnet(hrG.train.mtrx, hrG.training$HR, alpha = 0,
lambda = grid, thresh = 1e-12)

> hrG.ridge <- hrG.cv.ridge$lambda.min
    [1] 5.336699

> hrG.ridge.fit <- glmnet(hrG.train.mtrx, hrG.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrG.predict.ridge <- predict(hrG.ridge.fit, s = hrG.ridge,
    newx = hrG.test.mtrx)

> predict(hrG.ridge.fit, s = hrG.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 142.47276697
    (Intercept)   .
    Elevation     0.02702912
    Temp         -0.06954166
    Humidity      0.04314158
    DP           -0.02773350
    HI           -0.05905017
    Wind          0.09113976
    Gust          0.06203432
    Pressure      0.01811483

> mean((hrG.test$HR - hrG.predict.ridge)^2)
    [1] 289.7075

> sqrt(289.7075)
    [1] 17.0208
```

--

```
> hrH.cv.ridge <- cv.glmnet(hrH.train.mtrx, hrH.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrH.ridge <- hrH.cv.ridge$lambda.min
    [1] 0.869749

> hrH.ridge.fit <- glmnet(hrH.train.mtrx, hrH.training$HR, alpha = 0,
    lambda = grid, thresh = 1e-12)

> hrH.predict.ridge <- predict(hrH.ridge.fit, s = hrH.ridge,
    newx = hrH.test.mtrx)

> predict(hrH.ridge.fit, s = hrH.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 508.564577138
    (Intercept)   .
    Elevation     0.012665364
    Temp         -0.070423464
    Humidity      0.002236814
    DP           -0.021903169
    HI           -0.056105790
    Wind         -0.611789826
    Gust          0.083031335
    Pressure     -0.339439950
```

```
> mean((hrH.test$HR - hrH.predict.ridge)^2)
    [1] 167.2352

> sqrt(167.2352)
    [1] 12.93194
--
> paceA.cv.ridge <- cv.glmnet(paceA.train.mtrx, paceA.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)
> paceA.ridge <- paceA.cv.ridge$lambda.min
    [1] 0.01

> paceA.ridge.fit <- glmnet(paceA.train.mtrx, paceA.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceA.predict.ridge <- predict(paceA.ridge.fit, s = paceA.ridge,
    newx = paceA.test.mtrx)

> predict(paceA.ridge.fit, s = paceA.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 750.86692928
    (Intercept)   .
    Elevation     0.07805578
    Temp         -3.36957004
    Humidity      0.20231039
    DP           -1.36499675
    HI            4.85999011
    Wind         -0.51305405
    Gust         -0.25518163
    Pressure     -0.30368962

> mean((paceA.test$Pace - paceA.predict.ridge)^2)
    [1] 972.3125

> sqrt(972.3125)
    [1] 31.18193

> paceB.cv.ridge <- cv.glmnet(paceB.train.mtrx, paceB.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceB.ridge <- paceB.cv.ridge$lambda.min
    [1] 10.72267

> paceB.ridge.fit <- glmnet(paceB.train.mtrx, paceB.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceB.predict.ridge <- predict(paceB.ridge.fit, s = paceB.ridge,
    newx = paceB.test.mtrx)
```

```
> predict(paceB.ridge.fit, s = paceB.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 120.314402710
    (Intercept)   .
    Elevation     0.101736259
    Temp          0.269920304
    Humidity      0.005442533
    DP            0.264252443
    HI            0.241856183
    Wind         -0.383533247
    Gust          0.797742696
    Pressure      0.300343759

> mean((paceB.test$Pace - paceB.predict.ridge)^2)
    [1] 1348.605

> sqrt(1348.605)
    [1] 36.72336
--
> paceC.cv.ridge <- cv.glmnet(paceC.train.mtrx, paceC.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceC.ridge <- paceC.cv.ridge$lambda.min
    [1] 2.656088

> paceC.ridge.fit <- glmnet(paceC.train.mtrx, paceC.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceC.predict.ridge <- predict(paceC.ridge.fit, s = paceC.ridge,
    newx = paceC.test.mtrx)

> predict(paceC.ridge.fit, s = paceC.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 472.89401580
    (Intercept)   .
    Elevation     0.13551778
    Temp          0.10220647
    Humidity      0.06372625
    DP            0.12156300
    HI            0.10474776
    Wind          0.12407289
    Gust          0.35740327
    Pressure     -0.07480716

> mean((paceC.test$Pace - paceC.predict.ridge)^2)
    [1] 551.3664

> sqrt(551.3664)
    [1] 23.48119
--
> paceD.cv.ridge <- cv.glmnet(paceD.train.mtrx, paceD.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceD.ridge <- paceD.cv.ridge$lambda.min
    [1] 4.641589
```

```
> paceD.ridge.fit <- glmnet(paceD.train.mtrx, paceD.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceD.predict.ridge <- predict(paceD.ridge.fit, s = paceD.ridge,
    newx = paceD.test.mtrx)

> predict(paceD.ridge.fit, s = paceD.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 869.21069998
    (Intercept)   .
    Elevation    -0.02940814
    Temp          0.24644380
    Humidity     -0.22747127
    DP            0.10227365
    HI            0.23631715
    Wind         -1.41154788
    Gust         -0.37891611
    Pressure     -0.38297199

> mean((paceD.test$Pace - paceD.predict.ridge)^2)
    [1] 1647.953

> sqrt(1647.953)
    [1] 40.59499
```

--

```
> paceE.cv.ridge <- cv.glmnet(paceE.train.mtrx, paceE.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceE.ridge <- paceE.cv.ridge$lambda.min
    [1] 0.04037017

> paceE.ridge.fit <- glmnet(paceE.train.mtrx, paceE.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceE.predict.ridge <- predict(paceE.ridge.fit, s = paceE.ridge,
    newx = paceE.test.mtrx)

> predict(paceE.ridge.fit, s = paceE.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 805.53681679
    (Intercept)   .
    Elevation     0.04299048
    Temp         -1.13698347
    Humidity     -0.79584214
    DP            1.73341727
    HI           -0.37594924
    Wind         -0.26099468
    Gust          0.41513678
    Pressure     -0.32855150

> mean((paceE.test$Pace - paceE.predict.ridge)^2)
    [1] 1113.109

> sqrt(1113.109)
    [1] 33.36329
```

--

```
> paceF.cv.ridge <- cv.glmnet(paceF.train.mtrx, paceF.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceF.ridge <- paceF.cv.ridge$lambda.min
    [1] 12.32847

> paceF.ridge.fit <- glmnet(paceF.train.mtrx, paceF.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceF.predict.ridge <- predict(paceF.ridge.fit, s = paceF.ridge,
    newx = paceF.test.mtrx)

> predict(paceF.ridge.fit, s = paceF.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 249.583354928
    (Intercept)   .
    Elevation     0.079948541
    Temp          0.304293584
    Humidity     -0.188544151
    DP            0.106225906
    HI            0.242108478
    Wind         -0.126392748
    Gust          0.003618072
    Pressure      0.135358654

> mean((paceF.test$Pace - paceF.predict.ridge)^2)
    [1] 1053.876

> sqrt(1053.876)
    [1] 32.46346
--
> paceG.cv.ridge <- cv.glmnet(paceG.train.mtrx, paceG.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceG.ridge <- paceG.cv.ridge$lambda.min
    [1] 0.01

> paceG.ridge.fit <- glmnet(paceG.train.mtrx, paceG.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceG.predict.ridge <- predict(paceG.ridge.fit, s = paceG.ridge,
    newx = paceG.test.mtrx)

> predict(paceG.ridge.fit, s = paceG.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 302.30954057
    (Intercept)   .
    Elevation     0.08756479
    Temp          6.73529550
    Humidity     -0.58508897
    DP            1.25030326
    HI           -6.46996985
    Wind          0.07224814
    Gust         -0.15674585
    Pressure      0.04004814
```

```
> mean((paceG.test$Pace - paceG.predict.ridge)^2)
    [1] 1262.433

> sqrt(1262.433)
    [1] 35.53073
--
> paceH.cv.ridge <- cv.glmnet(paceH.train.mtrx, paceH.training$Pace,
    alpha = 0, lambda = grid, thresh = 1e-12)

> paceH.ridge <- paceH.cv.ridge$lambda.min
    [1] 49.77024

> paceH.ridge.fit <- glmnet(paceH.train.mtrx, paceH.training$Pace,
     alpha = 0, lambda = grid, thresh = 1e-12)

> paceH.predict.ridge <- predict(paceH.ridge.fit, s = paceH.ridge,
    newx = paceH.test.mtrx)

> predict(paceH.ridge.fit, s = paceH.ridge, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 277.617297074
    (Intercept)   .
    Elevation     0.006737611
    Temp          0.042250340
    Humidity     -0.076133051
    DP           -0.011845568
    HI            0.026111492
    Wind          0.705604133
    Gust          0.118468396
    Pressure      0.113279560

> mean((paceH.test$Pace - paceH.predict.ridge)^2)
    [1] 2798.588

> sqrt(2798.588)
    [1] 52.90168
```

Lasso

```
> hr.cv.lasso <- cv.glmnet(hr.train.mtrx, hr.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hr.lasso <- hr.cv.lasso$lambda.min
    [1] 0.01

> hr.lasso.fit <- glmnet(train.mtrx, training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hr.predict.lasso <- predict(hr.lasso.fit, s = hr.lasso,
    newx = test.mtrx)
```

```
> predict(hr.lasso.fit, s = hr.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 138.219035946
    (Intercept)   .
    Elevation     0.002495752
    Temp          0.202831920
    Humidity      0.222664093
    DP           -0.343477580
    HI            0.002253930
    Wind         -0.021831449
    Gust         -0.012322573
    Pressure      0.006768119

> mean((test$HR - hr.predict.lasso)^2)
    [1] 217.874

> sqrt(217.874)
    [1] 14.76056
--
> pace.cv.lasso <- cv.glmnet(pace.train.mtrx, pace.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> pace.lasso <- pace.cv.lasso$lambda.min
    [1] 0.01

> pace.lasso.fit <- glmnet(pace.train.mtrx, pace.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> pace.predict.lasso <- predict(pace.lasso.fit, s = pace.lasso,
    newx = pace.test.mtrx)

> predict(pace.lasso.fit, s = pace.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 246.41359530
    (Intercept)   .
    Elevation     0.07834315
    Temp          4.12818623
    Humidity      0.87493772
    DP           -0.98418747
    HI           -2.76324596
    Wind          0.01110645
    Gust         -0.03179243
    Pressure      0.08012099

> mean((pace.test$Pace - pace.predict.lasso)^2)
    [1] 2314.781

> sqrt(2314.781)
    [1] 48.11217
```

It is very peculiar to note that the lasso regression has not reduced any of the predictors to zero and is essentially acting similar to a ridge regression for the dataset as whole. It can be seen below that the lasso does reduce some of the predictor values to zero for the gender and individual subsets. However as seen in previous analysis, the relevance of these subsets and their significance in being able to offer a method predicting response values for the performance metrics as response variables have been questionable at best.

```
> hrW.cv.lasso <- cv.glmnet(hrW.train.mtrx, hrW.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrW.lasso <- hrW.cv.lasso$lambda.min
    [1] 0.07054802

> hrW.lasso.fit <- glmnet(train.mtrx, training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrW.lasso.fit <- glmnet(hrW.train.mtrx, hrW.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrW.predict.lasso <- predict(hrW.lasso.fit, s = hrW.lasso,
    newx = hrW.test.mtrx)

> predict(hrW.lasso.fit, s = hrW.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 130.71486258
    (Intercept)   .
    Elevation    -0.01071675
    Temp          0.06843362
    Humidity      0.07066790
    DP            .
    HI            .
    Wind          0.18047417
    Gust          .
    Pressure      0.01893494

> mean((hrW.test$HR - hrW.predict.lasso)^2)
    [1] 167.9064

> sqrt(167.9064)
    [1] 12.95787
```

--

```
> hrM.cv.lasso <- cv.glmnet(hrM.train.mtrx, hrM.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrM.lasso <- hrM.cv.lasso$lambda.min
    [1] 0.3274549

> hrM.lasso.fit <- glmnet(hrM.train.mtrx, hrM.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrM.predict.lasso <- predict(hrM.lasso.fit, s = hrM.lasso,
    newx = hrM.test.mtrx)
```

```
> predict(hrM.lasso.fit, s = hrM.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 161.478717565
    (Intercept)   .
    Elevation     .
    Temp         -0.151367040
    Humidity      0.006289439
    DP            .
    HI            .
    Wind          .
    Gust          .
    Pressure      .

> mean((hrM.test$HR - hrM.predict.lasso)^2)
    [1] 227.6411

> sqrt(227.6411)
    [1] 15.08778

--

> hrA.cv.lasso <- cv.glmnet(hrA.train.mtrx, hrA.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrA.lasso <- hrA.cv.lasso$lambda.min
    [1] 0.3274549

> hrA.lasso.fit <- glmnet(hrA.train.mtrx, hrA.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrA.predict.lasso <- predict(hrA.lasso.fit, s = hrA.lasso,
    newx = hrA.test.mtrx)

> predict(hrA.lasso.fit, s = hrA.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 179.94477616
    (Intercept)   .
    Elevation     .
    Temp          0.09671183
    Humidity      .
    DP            .
    HI            .
    Wind          .
    Gust          0.14007809
    Pressure     -0.02579427

> mean((hrA.test$HR - hrA.predict.lasso)^2)
    [1] 132.7447

> sqrt(132.7447)
    [1] 11.52149

--

> hrB.cv.lasso <- cv.glmnet(hrB.train.mtrx, hrB.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrB.lasso <- hrB.cv.lasso$lambda.min
    [1] 10000
```

```
> hrB.lasso.fit <- glmnet(hrB.train.mtrx, hrB.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrB.predict.lasso <- predict(hrB.lasso.fit, s = hrB.lasso,
    newx = hrB.test.mtrx)

> predict(hrB.lasso.fit, s = hrB.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                        s1
    (Intercept) 160.6021
    (Intercept)   .
    Elevation     .
    Temp          .
    Humidity      .
    DP            .
    HI            .
    Wind          .
    Gust          .
    Pressure      .

> mean((hrB.test$HR - hrB.predict.lasso)^2)
    [1] 137.6317

> sqrt(137.6317)
    [1] 11.73165
--
> hrC.cv.lasso <- cv.glmnet(hrC.train.mtrx, hrC.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrC.lasso <- hrC.cv.lasso$lambda.min
    [1] 0.01

> hrC.lasso.fit <- glmnet(hrC.train.mtrx, hrC.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrC.predict.lasso <- predict(hrC.lasso.fit, s = hrC.lasso,
    newx = hrC.test.mtrx)

> predict(hrC.lasso.fit, s = hrC.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 163.91013869
    (Intercept)   .
    Elevation    -0.01392971
    Temp          0.28075779
    Humidity      0.14211101
    DP           -0.28471438
    HI            .
    Wind          0.18937822
    Gust         -0.22704005
    Pressure     -0.02376169

> mean((hrC.test$HR - hrC.predict.lasso)^2)
    [1] 103.7852

> sqrt(103.7852)
    [1] 10.1875
```

--

```
> hrD.cv.lasso <- cv.glmnet(hrD.train.mtrx, hrD.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrD.lasso <- hrD.cv.lasso$lambda.min
    [1] 0.3274549

> hrD.lasso.fit <- glmnet(hrD.train.mtrx, hrD.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrD.predict.lasso <- predict(hrD.lasso.fit, s = hrD.lasso,
    newx = hrD.test.mtrx)

> predict(hrD.lasso.fit, s = hrD.lasso, type = "coefficients")
    10 x 1 sparse Matrix of class "dgCMatrix"
                         s1
    (Intercept) 354.8247213
    (Intercept)   .
    Elevation     .
    Temp          .
    Humidity      .
    DP            0.1437681
    HI            .
    Wind          0.1852943
    Gust          .
    Pressure     -0.1894402

> mean((hrD.test$HR - hrD.predict.lasso)^2)
    [1] 150.8564

> sqrt(150.8564)
    [1] 12.28236
```

--

```
> hrE.cv.lasso <- cv.glmnet(hrE.train.mtrx, hrE.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrE.lasso <- hrE.cv.lasso$lambda.min
    [1] 0.6579332

> hrE.lasso.fit <- glmnet(hrE.train.mtrx, hrE.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrE.predict.lasso <- predict(hrE.lasso.fit, s = hrE.lasso,
    newx = hrE.test.mtrx)

> predict(hrE.lasso.fit, s = hrE.lasso, type = "coefficients")
    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 151.41694445
    (Intercept)   .
    Elevation     .
    Temp          .
    Humidity      .
    DP           -0.05110689
    HI           -0.04326714
    Wind          .
    Gust          .
    Pressure      .
```

```
> mean((hrE.test$HR - hrE.predict.lasso)^2)
    [1] 180.3651

> sqrt(180.3651)
    [1] 13.43001
```

--

```
> hrF.cv.lasso <- cv.glmnet(hrF.train.mtrx, hrF.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrF.lasso <- hrF.cv.lasso$lambda.min
    [1] 0.2477076

> hrF.lasso.fit <- glmnet(hrF.train.mtrx, hrF.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrF.predict.lasso <- predict(hrF.lasso.fit, s = hrF.lasso,
    newx = hrF.test.mtrx)

> predict(hrF.lasso.fit, s = hrF.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 242.90388893
    (Intercept)   .
    Elevation     .
    Temp          .
    Humidity      0.05146504
    DP            .
    HI           -0.21138030
    Wind          .
    Gust          .
    Pressure     -0.07498478

> mean((hrF.test$HR - hrF.predict.lasso)^2)
    [1] 149.856

> sqrt(149.856)
    [1] 12.24157
```

--

```
> hrG.cv.lasso <- cv.glmnet(hrG.train.mtrx, hrG.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrG.lasso <- hrG.cv.lasso$lambda.min
    [1] 0.1072267

> hrG.lasso.fit <- glmnet(hrG.train.mtrx, hrG.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrG.predict.lasso <- predict(hrG.lasso.fit, s = hrG.lasso,
    newx = hrG.test.mtrx)
```

```
> predict(hrG.lasso.fit, s = hrG.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 143.87982174
    (Intercept)   .
    Elevation     0.03174868
    Temp         -0.17458730
    Humidity      0.03380165
    DP            .
    HI            .
    Wind          0.10575918
    Gust          0.06788681
    Pressure      0.01888011

> mean((hrG.test$HR - hrG.predict.lasso)^2)
    [1] 289.619

> sqrt(289.619)
    [1] 17.0182
--
> hrH.cv.lasso <- cv.glmnet(hrH.train.mtrx, hrH.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrH.lasso <- hrH.cv.lasso$lambda.min
    [1] 0.03053856

> hrH.lasso.fit <- glmnet(hrH.train.mtrx, hrH.training$HR, alpha = 1,
    lambda = grid, thresh = 1e-12)

> hrH.predict.lasso <- predict(hrH.lasso.fit, s = hrH.lasso,
    newx = hrH.test.mtrx)

> predict(hrH.lasso.fit, s = hrH.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 527.42898360
    (Intercept)   .
    Elevation     0.01300626
    Temp         -0.15538361
    Humidity     -0.01420487
    DP            .
    HI            .
    Wind         -0.65687412
    Gust          0.10077501
    Pressure     -0.35604102

> mean((hrH.test$HR - hrH.predict.lasso)^2)
    [1] 168.0663

> sqrt(168.0663)
    [1] 12.96404
--
> paceM.cv.lasso <- cv.glmnet(paceM.train.mtrx, paceM.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceM.lasso <- paceM.cv.lasso$lambda.min
    [1] 0.01321941
```

```
> paceM.lasso.fit <- glmnet(paceM.train.mtrx, paceM.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceM.predict.lasso <- predict(paceM.lasso.fit, s = paceM.lasso,
    newx = paceM.test.mtrx)

> predict(paceM.lasso.fit, s = paceM.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 318.329123982
    (Intercept)   .
    Elevation     0.093397865
    Temp          7.645770058
    Humidity      0.177394433
    DP            0.151911175
    HI           -6.643631437
    Wind          0.149266621
    Gust         -0.014007839
    Pressure     -0.004803381

> mean((paceM.test$Pace - paceM.predict.lasso)^2)
    [1] 1583.367

> sqrt(1583.367)
    [1] 39.79154
```

--

```
> paceA.cv.lasso <- cv.glmnet(paceA.train.mtrx, paceA.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceA.lasso <- paceA.cv.lasso$lambda.min
    [1] 0.01321941

> paceA.lasso.fit <- glmnet(paceA.train.mtrx, paceA.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceA.predict.lasso <- predict(paceA.lasso.fit, s = paceA.lasso,
    newx = paceA.test.mtrx)

> predict(paceA.lasso.fit, s = paceA.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 776.88529866
    (Intercept)   .
    Elevation     0.07765117
    Temp         -5.58506774
    Humidity      0.04166453
    DP           -1.19010555
    HI            6.71807225
    Wind         -0.49219390
    Gust         -0.27550266
    Pressure     -0.30384607

> mean((paceA.test$Pace - paceA.predict.lasso)^2)
    [1] 973.7255

> sqrt(973.7255)
    [1] 31.20457
```

--

```
> paceB.cv.lasso <- cv.glmnet(paceB.train.mtrx, paceB.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceB.lasso <- paceB.cv.lasso$lambda.min
    [1] 0.2477076

> paceB.lasso.fit <- glmnet(paceB.train.mtrx, paceB.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceB.predict.lasso <- predict(paceB.lasso.fit, s = paceB.lasso,
    newx = paceB.test.mtrx)

> predict(paceB.lasso.fit, s = paceB.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                         s1
    (Intercept) 63.1784703
    (Intercept)  .
    Elevation    0.1228070
    Temp         0.4211537
    Humidity     .
    DP           0.4144562
    HI           .
    Wind        -0.4152641
    Gust         0.9935091
    Pressure     0.3548384

> mean((paceB.test$Pace - paceB.predict.lasso)^2)
    [1] 1370.198

> sqrt(1370.198)
    [1] 37.01619
--
> paceC.cv.lasso <- cv.glmnet(paceC.train.mtrx, paceC.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceC.lasso <- paceC.cv.lasso$lambda.min
    [1] 0.01321941

> paceC.lasso.fit <- glmnet(paceC.train.mtrx, paceC.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceC.predict.lasso <- predict(paceC.lasso.fit, s = paceC.lasso,
    newx = paceC.test.mtrx)

> predict(paceC.lasso.fit, s = paceC.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 485.17682929
    (Intercept)   .
    Elevation     0.14845888
    Temp          .
    Humidity      0.10043440
    DP            0.07157824
    HI            0.25334605
    Wind          0.14758964
    Gust          0.39152408
    Pressure     -0.09019720
```

```
> mean((paceC.test$Pace - paceC.predict.lasso)^2)
    [1] 553.4324

> sqrt(553.4324)
    [1] 23.52514
```

--

```
> paceD.cv.lasso <- cv.glmnet(paceD.train.mtrx, paceD.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceD.lasso <- paceD.cv.lasso$lambda.min
    [1] 0.2477076

> paceD.lasso.fit <- glmnet(paceD.train.mtrx, paceD.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceD.predict.lasso <- predict(paceD.lasso.fit, s = paceD.lasso,
    newx = paceD.test.mtrx)

> predict(paceD.lasso.fit, s = paceD.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 919.26012950
    (Intercept)   .
    Elevation    -0.02201928
    Temp          .
    Humidity     -0.22205921
    DP            .
    HI            0.54794479
    Wind         -1.58472581
    Gust         -0.31859359
    Pressure     -0.42996465

> mean((paceD.test$Pace - paceD.predict.lasso)^2)
    [1] 1645.333

> sqrt(1645.333)
    [1] 40.5627
```

--

```
> paceE.cv.lasso <- cv.glmnet(paceE.train.mtrx, paceE.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceE.lasso <- paceE.cv.lasso$lambda.min
    [1] 0.01

> paceE.lasso.fit <- glmnet(paceE.train.mtrx, paceE.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceE.predict.lasso <- predict(paceE.lasso.fit, s = paceE.lasso,
    newx = paceE.test.mtrx)
```

```
> predict(paceE.lasso.fit, s = paceE.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 828.45232571
    (Intercept)   .
    Elevation     0.04245024
    Temp         -1.76345648
    Humidity     -0.91645924
    DP            1.95716733
    HI            .
    Wind         -0.26203691
    Gust          0.41076614
    Pressure     -0.33849303

> mean((paceE.test$Pace - paceE.predict.lasso)^2)
    [1] 1116.213

> sqrt(1116.213)
    [1] 33.40977
--
> paceF.cv.lasso <- cv.glmnet(paceF.train.mtrx, paceF.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceF.lasso <- paceF.cv.lasso$lambda.min
    [1] 1

> paceF.lasso.fit <- glmnet(paceF.train.mtrx, paceF.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceF.predict.lasso <- predict(paceF.lasso.fit, s = paceF.lasso,
    newx = paceF.test.mtrx)

> predict(paceF.lasso.fit, s = paceF.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                           s1
    (Intercept) 311.16964337
    (Intercept)   .
    Elevation     0.05964315
    Temp          0.68903522
    Humidity     -0.13159235
    DP            .
    HI            .
    Wind          .
    Gust          .
    Pressure      0.06624537

> mean((paceF.test$Pace - paceF.predict.lasso)^2)
    [1] 1054.789

> sqrt(1054.789)
    [1] 32.47752
--
> paceG.cv.lasso <- cv.glmnet(paceG.train.mtrx, paceG.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceG.lasso <- paceG.cv.lasso$lambda.min
    [1] 0.02009233
```

```
> paceG.lasso.fit <- glmnet(paceG.train.mtrx, paceG.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceG.predict.lasso <- predict(paceG.lasso.fit, s = paceG.lasso,
    newx = paceG.test.mtrx)

> predict(paceG.lasso.fit, s = paceG.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 275.31547634
    (Intercept)   .
    Elevation     0.08699457
    Temp          9.14606490
    Humidity     -0.42722144
    DP            1.13357039
    HI           -8.56370430
    Wind          0.07990609
    Gust         -0.15746935
    Pressure      0.04062621

> mean((paceG.test$Pace - paceG.predict.lasso)^2)
    [1] 1262.535

> sqrt(1262.535)
    [1] 35.53217
--
> paceH.cv.lasso <- cv.glmnet(paceH.train.mtrx, paceH.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)*

> paceH.lasso <- paceH.cv.lasso$lambda.min
    [1] 2.009233

> paceH.lasso.fit <- glmnet(paceH.train.mtrx, paceH.training$Pace,
    alpha = 1, lambda = grid, thresh = 1e-12)

> paceH.predict.lasso <- predict(paceH.lasso.fit, s = paceH.lasso,
    newx = paceH.test.mtrx)

> predict(paceH.lasso.fit, s = paceH.lasso, type = "coefficients")

    10 x 1 sparse Matrix of class "dgCMatrix"
                          s1
    (Intercept) 342.39702703
    (Intercept)   .
    Elevation     .
    Temp          .
    Humidity     -0.04549037
    DP            .
    HI            .
    Wind          1.15282565
    Gust          .
    Pressure      0.04703922

> mean((paceH.test$Pace - paceH.predict.lasso)^2)
    [1] 2798.512

> sqrt(2798.512)
    [1] 52.9
```

Each of the predictive analyses used have produced similar error testing values in their validations. For this reason it is suggested that the Best Subset Selection model should be used in attempting to calculate response for the performance based metrics HR and Pace. This model is the only one that successfully reduced the number predictor values it used in its prediction and these values (for the dataset as a whole) matched the predictors that showed significant p-values when tested in a linear regression analysis.

Summary

Findings

Firstly, it was surprising not to find greater levels of correlation between the performance metrics HR and Pace in respect to the weather based metrics namely, Temp, Humidity, DP and HI. The only significant correlation values produced from the dataset of variables were between the weather metrics which are all derived from each other in their calculations and so are correlated anyway. The correlation plots then offered little further insight into the dataset and its relationships. The only evident point of note was that there were clearly a lot of zero values under the Wind and Gust attribute headers.

It was again surprising to find that the plots of HR and Pace were not normally distributed despite appearing so. The small population size may have played a role in this discovery or the small number of instances that were seen to exist at the outer extremes of the range.

The sctterplots produced did little to answer the question of whether the time of day at which an activity was conducted affected the HR or Pace values that were being measured. They did however back up the cited limitations of the data that was expected, in that athletes are more likely to workout at earlier hours of the morning in order to avoid the harsh conditions of the environment in the midday hours. This was evident through the greater number of points corresponding to earlier hours of the day preceding 09:00.

The density graphs produced, offered an interesting visual of the average HR and Pace values in respect to the DP values recorded during an activity. There was a slight suggestion from these visualisations that the higher DP values aligned with higher HR values, in general it appeared to show nominal HR and Pace values.

Clustering proved to be less than fruitful when it comes to attempting to understand if the different ranges in DewFeel align with specific values of HR and Pace. The plots produced were a mess of colour suggesting that there is in fact great variability and distribution within the data that was sampled.

The boxplots didn’t show any large differences in the performance metrics, HR and Pace between each of the DewFeel categories. The differences were subtle, unfortunately not boding well of the rest of the analysis or the overarching hypothesis in general.

The contingency table generated allowed the observer to see the distribution of each individual's activities across the whole humidity range. It was clear that each individual had spent a large amount of time conducting activities across each category which would allow for greater understanding of the effects of higher humidity on performance. The heatmaps allowed for this numerical data to be visualised, which made it apparent that each individual spent most of their time in dry environments primarily due to the large range of degrees the category encompasses.

The simple linear regression models supported the expectation that DP and HI would prove to be influential metrics when it came to predicting the response from the performance metrics of HR and Pace. This cause effect relationship was evident even from the gender and individual subsets.

The multiple regression models supported the belief that there was likely a synergistic relationship between the performance metrics HR and Pace and the primary weather predictors DP and HI. It was interesting to see that HR and Pace had little ability to predict what DP it might be during an activity but in hindsight this was unlikely largely due to the variability noticed in the data.

The predictive analyses shown in this report were concerning. Each one of them suggested significance from the p-values produced for the predictor values and F-statistics, but the r-squared values and error testing lended little credibility to the models. It is suspected that there likely is a measurable effect from the weather metrics on the performance metrics but the relatively small size of the dataset in conjunction with the evident variability in the data itself made it difficult to conclusively generate any models of statistical merit.

Weaknesses

It is immediately clear that the size of the dataset used hindered the statistical relevance of this analysis. A much larger dataset would likely have done a much better job at filtering out the straying data values that appear to have been causing the variability in the data to be misaligned.

In a broader sense the collection of the data for this analysis proved to be a monumental challenge. It proved exceedingly difficult to scrap large swaths of data from the available sources and so much of the collection process involved manual transcription from individual web pages into Excel. Had it been easier to access the raw data a much larger dataset comprising many more individuals could have been compiled offering greater insight from this analysis.

Finally the relative newness of the performance measuring features provided in fitness trackers was a small hurdle. Many of the original data subjects that had hoped to be sampled unfortunately did not have access to devices that recorded HR values while they were based in humid regions. The technology has on recently become standard and so data earlier than 2017 is very difficult to find.

Future

First off, as has already been mentioned, a much larger dataset is required in order to offer any substantial insight/evidence to support or reject the hypothesis of the report. Subsequently, it will be interesting to continue this research using a variety of different performance based metrics as these are the metrics that will offer the greatest insight versus the well understood weather data. Wrist based devices have in very recent times begun to add ever more advanced features such as blood oxygen measurements, ECG measurements, perceived effort predictors etc. Further to this using some slightly more invasive methods such as blood lactate level testing, air/oxygen intake, even understanding subject water weight contents before and after exercise to gauge sweat loss etc will all add so much nuance to the analysis of the effect of humidity on high performance athletics. Many of these very metrics were discussed at the outset of this project, but due to the sheer scale of possibility it was decided to err on the side of caution and begin a simple exploratory analysis to understand if evidence of the hypothesis posed can be seen at the lowest most grassroots form of performance measuring.

Appendix

Formula to calculate Heat Index (HI)

```
HI = 0.5 * (Temp + 61 + ((Temp-68)*1.2) + (Humidity*0.094))
```

Function used to calculate DewFeel

```
DewFeel = IF(DP<50, "Dry", IF(DP<56, "Pleasant", IF(DP<61,
"Comfortable", IF(DP<66, "Sticky", IF(DP<71, "Uncomfortable",
IF(DP<76,"Oppressive","Miserable"))))))
```

Dew Point	Comfort Level
< 50	Dry
50 - 55	Pleasant
56-60	Comfortable
61-65	Sticky
66-70	Uncomfortable
71-75	Oppressive
76 +	Miserable

Functions used to set up predictive analyses

```
> hrW.predictors <- gender$F[c(7:12, 15:17)]
> paceW.predictors <- gender$F[c(6, 8:12, 15:17)]

> hrM.predictors <- gender$M[c(7:12, 15:17)]
> paceM.predictors <- gender$M[c(6, 8:12, 15:17)]

> hrA.predictors <- subject$A[c(7:12, 15:17)]
> paceA.predictors <- subject$A[c(6, 8:12, 15:17)]

> hrB.predictors <- subject$B[c(7:12, 15:17)]
> paceB.predictors <- subject$B[c(6, 8:12, 15:17)]

> hrC.predictors <- subject$C[c(7:12, 15:17)]
> paceC.predictors <- subject$C[c(6, 8:12, 15:17)]

> hrD.predictors <- subject$D[c(7:12, 15:17)]
> paceD.predictors <- subject$D[c(6, 8:12, 15:17)]

> hrD.predictors <- subject$E[c(7:12, 15:17)]
> paceD.predictors <- subject$E[c(6, 8:12, 15:17)]

> hrE.predictors <- subject$F[c(7:12, 15:17)]
> paceE.predictors <- subject$F[c(6, 8:12, 15:17)]

> hrF.predictors <- subject$G[c(7:12, 15:17)]
> paceF.predictors <- subject$G[c(6, 8:12, 15:17)]

> hrG.predictors <- subject$H[c(7:12, 15:17)]
> paceG.predictors <- subject$H[c(6, 8:12, 15:17)]
```

--

```
> hrW.part <- createDataPartition(y = hrW.predictors$HR, p = 0.7,
      list=FALSE)

> hrW.training <- hrW.predictors[hrW.part, ]
> hrW.test <- hrW.predictors[-hrW.part, ]
```

```
   > paceW.training <- paceW.predictors[paceW.part, ]
   > paceW.test <- paceW.predictors[-paceW.part, ]

> hrM.part <- createDataPartition(y = hrM.predictors$HR, p = 0.7,
       list=FALSE)

   > hrM.training <- hrM.predictors[hrM.part, ]
   > hrM.test <- hrM.predictors[-hrM.part, ]

   > paceM.training <- paceM.predictors[paceM.part, ]
   > paceM.test <- paceM.predictors[-paceM.part, ]

> hrA.part <- createDataPartition(y = hrA.predictors$HR, p = 0.7,
       list=FALSE)

   > hrA.training <- hrA.predictors[hrA.part, ]
   > hrA.test <- hrA.predictors[-hrA.part, ]

   > paceA.training <- paceA.predictors[paceA.part, ]
   > paceA.test <- paceA.predictors[-paceA.part, ]

> hrB.part <- createDataPartition(y = hrB.predictors$HR, p = 0.7,
       list=FALSE)

   > hrB.training <- hrB.predictors[hrB.part, ]
   > hrB.test <- hrB.predictors[-hrB.part, ]

   > paceB.training <- paceB.predictors[paceB.part, ]
   > paceB.test <- paceB.predictors[-paceB.part, ]

> hrC.part <- createDataPartition(y = hrC.predictors$HR, p = 0.7,
       list=FALSE)

   > hrC.training <- hrC.predictors[hrC.part, ]
   > hrC.test <- hrC.predictors[-hrC.part, ]

   > paceC.training <- paceC.predictors[paceC.part, ]
   > paceC.test <- paceC.predictors[-paceC.part, ]

> hrD.part <- createDataPartition(y = hrD.predictors$HR, p = 0.7,
       list=FALSE)

   > hrD.training <- hrD.predictors[hrD.part, ]
   > hrD.test <- hrD.predictors[-hrD.part, ]

   > paceD.training <- paceD.predictors[paceD.part, ]
   > paceD.test <- paceD.predictors[-paceD.part, ]

> hrE.part <- createDataPartition(y = hrE.predictors$HR, p = 0.7,
       list=FALSE)

   > hrE.training <- hrE.predictors[hrE.part, ]
   > hrE.test <- hrE.predictors[-hrE.part, ]

   > paceE.training <- paceE.predictors[paceE.part, ]
   > paceE.test <- paceE.predictors[-paceE.part, ]

> hrF.part <- createDataPartition(y = hrF.predictors$HR, p = 0.7,
       list=FALSE)

   > hrF.training <- hrF.predictors[hrF.part, ]
   > hrF.test <- hrF.predictors[-hrF.part, ]

   > paceF.training <- paceF.predictors[paceF.part, ]
   > paceF.test <- paceF.predictors[-paceF.part, ]

> hrG.part <- createDataPartition(y = hrG.predictors$HR, p = 0.7,
```

```
        list=FALSE)

   > hrG.training <- hrG.predictors[hrG.part, ]
   > hrG.test <- hrG.predictors[-hrG.part, ]

   > paceG.training <- paceG.predictors[paceG.part, ]
   > paceG.test <- paceG.predictors[-paceG.part, ]

> hrH.part <- createDataPartition(y = hrH.predictors$HR, p = 0.7,
        list=FALSE)

   > hrH.training <- hrH.predictors[hrH.part, ]
   > hrH.test <- hrH.predictors[-hrH.part, ]

   > paceH.training <- paceH.predictors[paceH.part, ]
   > paceH.test <- paceH.predictors[-paceH.part, ]

--

   > hrW.train.mtrx <- model.matrix(HR ~ ., data = hrW.training)
   > hrW.test.mtrx <- model.matrix(HR ~ ., data = hrW.test)

   > paceW.train.mtrx <- model.matrix(Pace ~ ., data = paceW.training)
   > paceW.test.mtrx <- model.matrix(Pace ~ ., data = paceW.test)

   > hrM.train.mtrx <- model.matrix(HR ~ ., data = hrM.training)
   > hrM.test.mtrx <- model.matrix(HR ~ ., data = hrM.test)

   > paceM.train.mtrx <- model.matrix(Pace ~ ., data = paceM.training)
   > paceM.test.mtrx <- model.matrix(Pace ~ ., data = paceM.test)

   > hrA.train.mtrx <- model.matrix(HR ~ ., data = hrA.training)
   > hrA.test.mtrx <- model.matrix(HR ~ ., data = hrA.test)

   > paceA.train.mtrx <- model.matrix(Pace ~ ., data = paceA.training)
   > paceA.test.mtrx <- model.matrix(Pace ~ ., data = paceA.test)

   > hrB.train.mtrx <- model.matrix(HR ~ ., data = hrB.training)
   > hrB.test.mtrx <- model.matrix(HR ~ ., data = hrB.test)

   > paceB.train.mtrx <- model.matrix(Pace ~ ., data = paceB.training)
   > paceB.test.mtrx <- model.matrix(Pace ~ ., data = paceB.test)

   > hrC.train.mtrx <- model.matrix(HR ~ ., data = hrC.training)
   > hrC.test.mtrx <- model.matrix(HR ~ ., data = hrC.test)

   > paceC.train.mtrx <- model.matrix(Pace ~ ., data = paceC.training)
   > paceC.test.mtrx <- model.matrix(Pace ~ ., data = paceC.test)

   > hrD.train.mtrx <- model.matrix(HR ~ ., data = hrD.training)
   > hrD.test.mtrx <- model.matrix(HR ~ ., data = hrD.test)

   > paceD.train.mtrx <- model.matrix(Pace ~ ., data = paceD.training)
   > paceD.test.mtrx <- model.matrix(Pace ~ ., data = paceD.test)

   > hrE.train.mtrx <- model.matrix(HR ~ ., data = hrE.training)
   > hrE.test.mtrx <- model.matrix(HR ~ ., data = hrE.test)

   > paceE.train.mtrx <- model.matrix(Pace ~ ., data = paceE.training)
   > paceE.test.mtrx <- model.matrix(Pace ~ ., data = paceE.test)

   > hrF.train.mtrx <- model.matrix(HR ~ ., data = hrF.training)
   > hrF.test.mtrx <- model.matrix(HR ~ ., data = hrF.test)

   > paceF.train.mtrx <- model.matrix(Pace ~ ., data = paceF.training)
   > paceF.test.mtrx <- model.matrix(Pace ~ ., data = paceF.test)

   > hrG.train.mtrx <- model.matrix(HR ~ ., data = hrG.training)
```

```
> hrG.test.mtrx <- model.matrix(HR ~ ., data = hrG.test)

> paceG.train.mtrx <- model.matrix(Pace ~ ., data = paceG.training)
> paceG.test.mtrx <- model.matrix(Pace ~ ., data = paceG.test)

> hrH.train.mtrx <- model.matrix(HR ~ ., data = hrH.training)
> hrH.test.mtrx <- model.matrix(HR ~ ., data = hrH.test)

> paceH.train.mtrx <- model.matrix(Pace ~ ., data = paceH.training)
> paceH.test.mtrx <- model.matrix(Pace ~ ., data = paceH.test)
```

*`cv.lasso` produced one or more warnings for lambda convergence

Eg.

```
Warning message:
from glmnet Fortran code (error code -99); Convergence for 99th lambda
value not reached after maxit=100000 iterations; solutions for larger
lambdas returned
```