431 Class 08

thomaselove.github.io/431

2020-09-17

Today's R Packages

```
library(NHANES)
library(janitor)
library(knitr)
library(broom)
library(magrittr)
library(patchwork)
library(tidyverse)

theme_set(theme_bw())
```

Later today (I hope) I'll also load the `rstanarm` package.

Learning By Example(s)

Six Simulated Data Sets

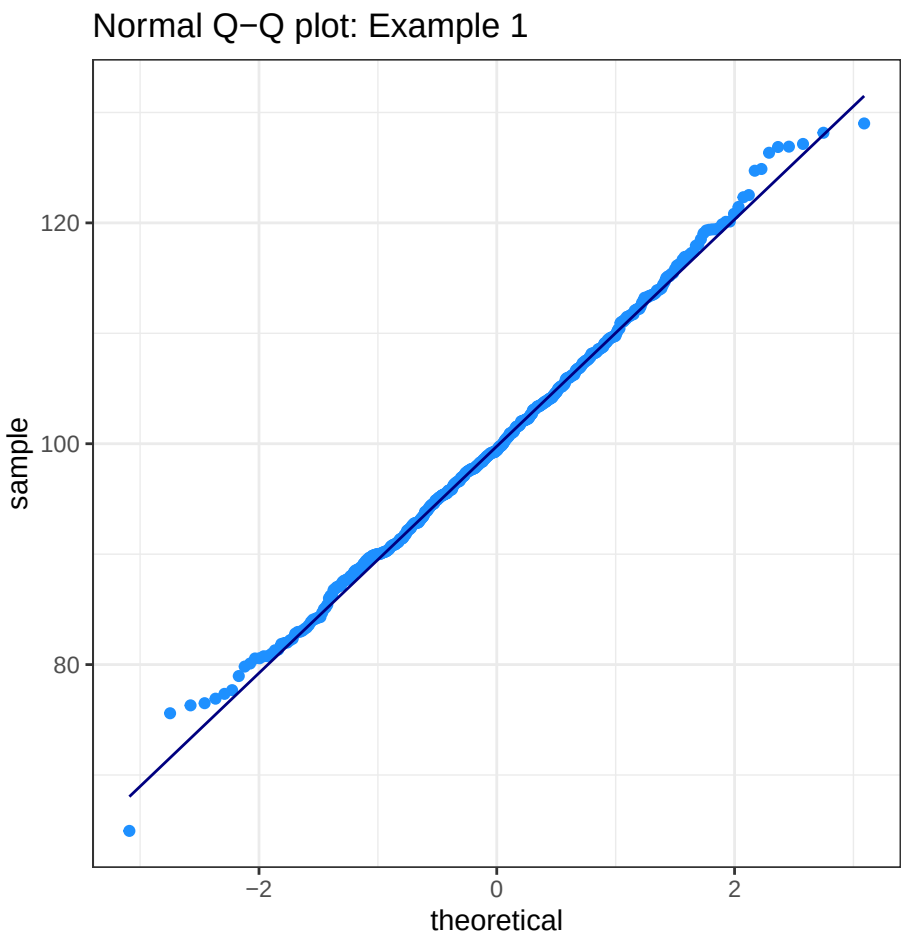
I built six example data sets, each containing 500 observations, and each drawing from a known distribution

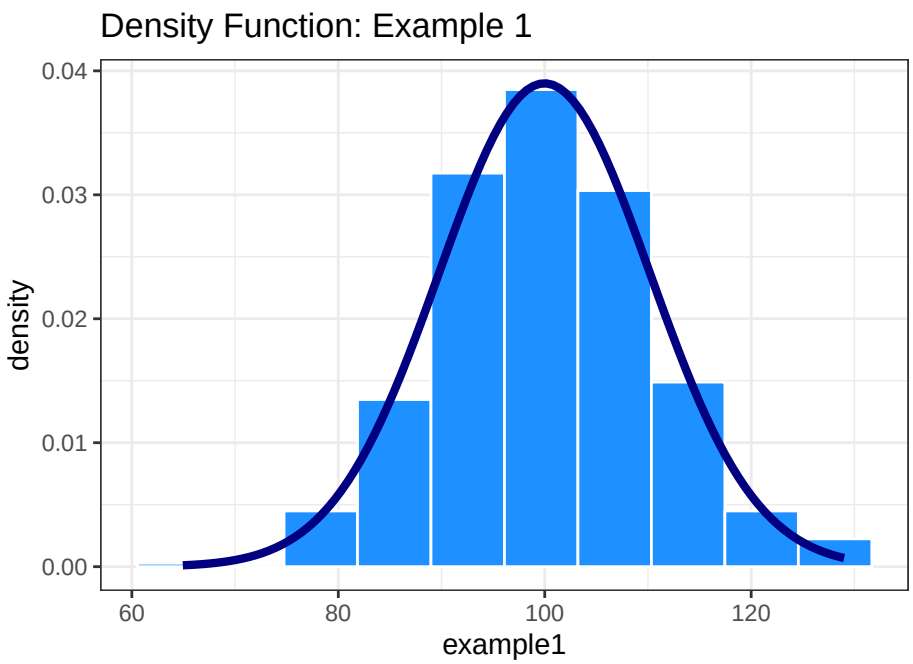
Set	Shape of Simulation Source
1	Normal distribution (symmetric, bell-shaped)
2	Left-skewed
3	Right-skewed
4	(Roughly) symmetric counts (integers)
5	Uniform on (0, 100) Symmetric, but flatter than a Normal
6	Symmetric, but with substantial outliers

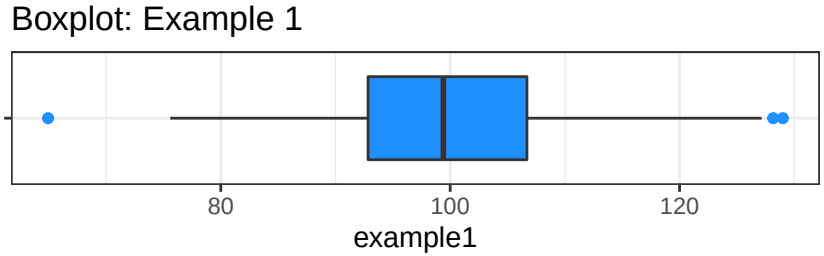
Does a Normal model fit well for these data?

1. Is a Normal Q-Q plot showing something close to a straight line, without clear signs of skew (as indicated by a curve in the plot) or indications of problems like outliers in the tails of the distribution (S-shape or reverse S-shape)?
2. Does a boxplot, violin plot and/or histogram show a symmetric distribution, where both the number of outliers is modest, and the distance of those outliers from the mean is modest?
3. Do numerical measures match up with the expectations of a normal model?

Example 1: Data from a Normal Distribution

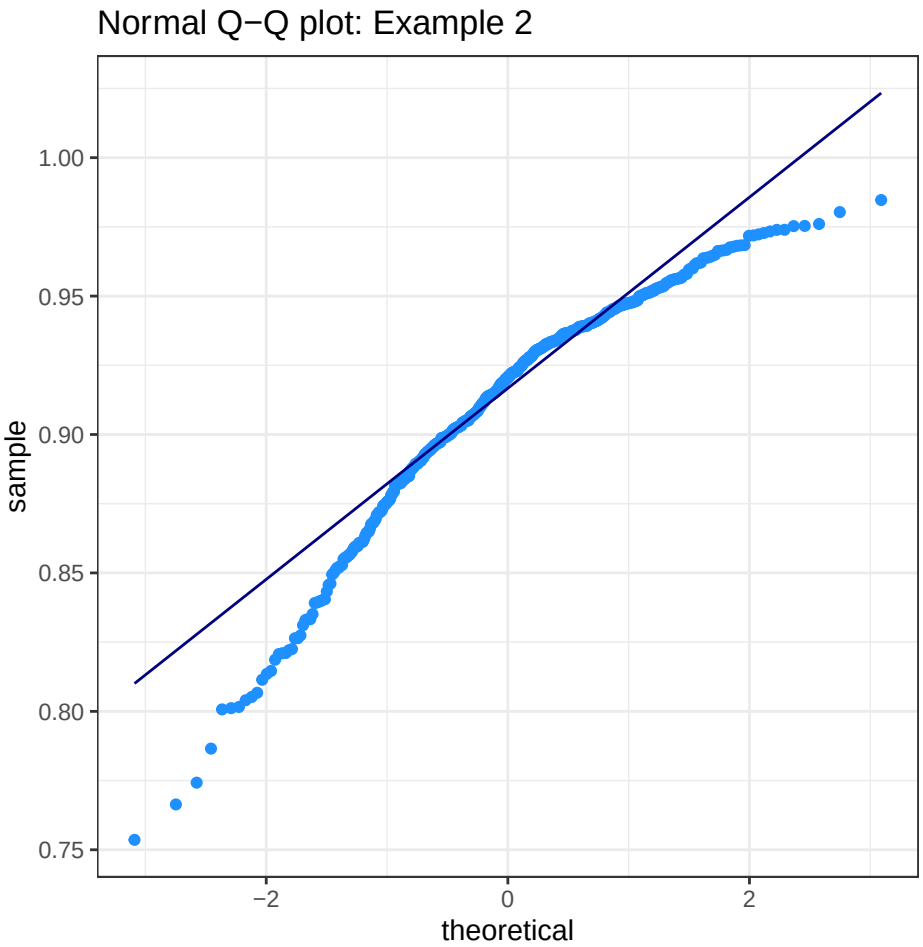


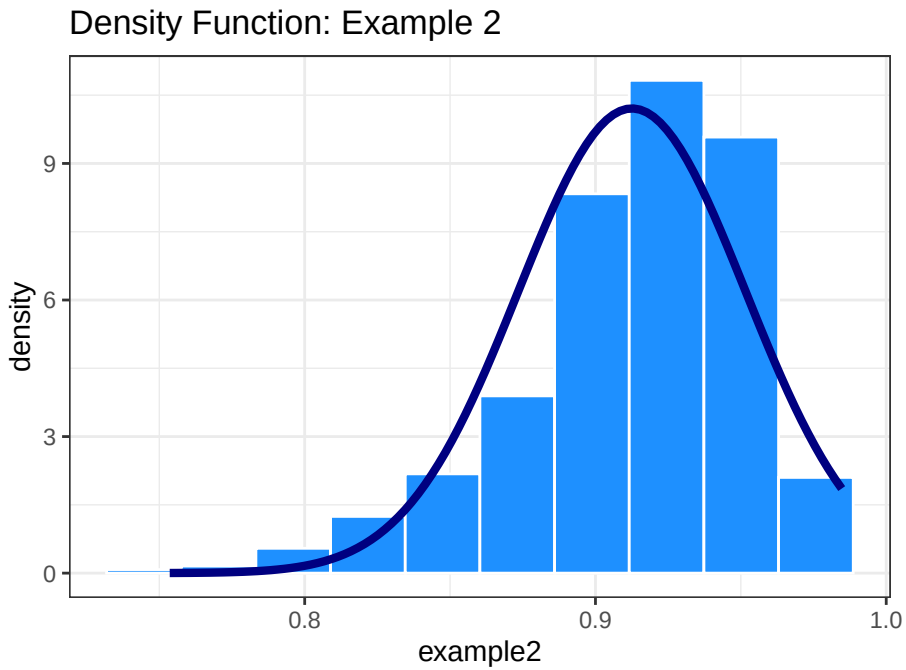


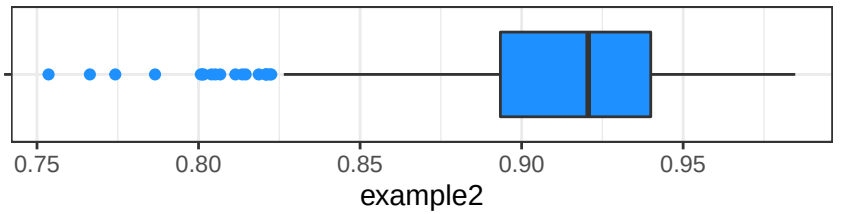


min	Q1	median	Q3	max	mean	sd	n	missing
64.9	92.8	99.4	106.7	129	100	10.2	500	0

Example 2: Data from a Left-Skewed Distribution

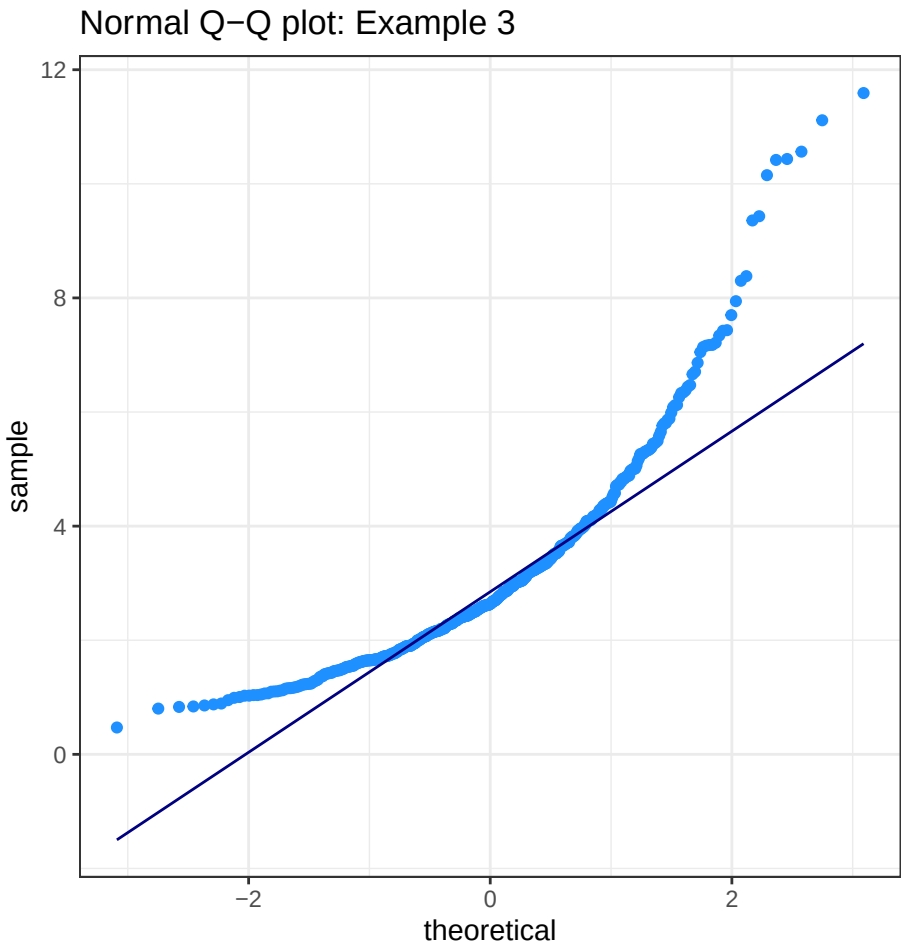


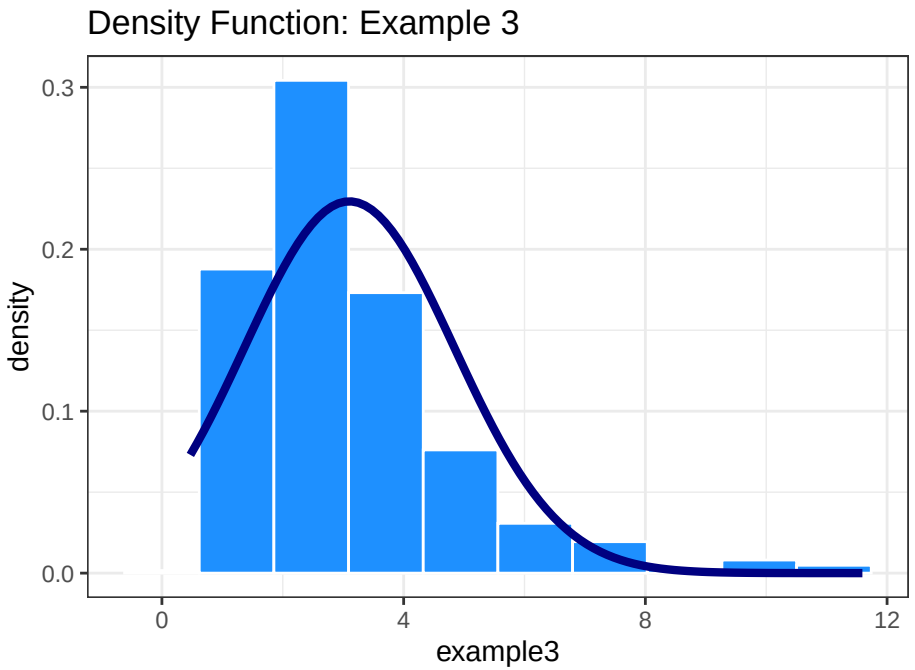


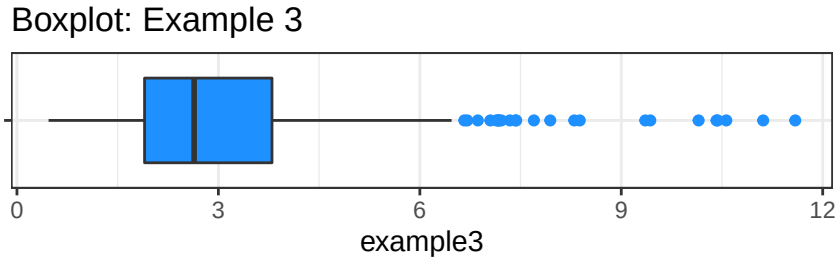


min	Q1	median	Q3	max	mean	sd	n	missing
0.75	0.89	0.92	0.94	0.98	0.91	0.04	500	0

Example 3: Data from a Right-Skewed Distribution

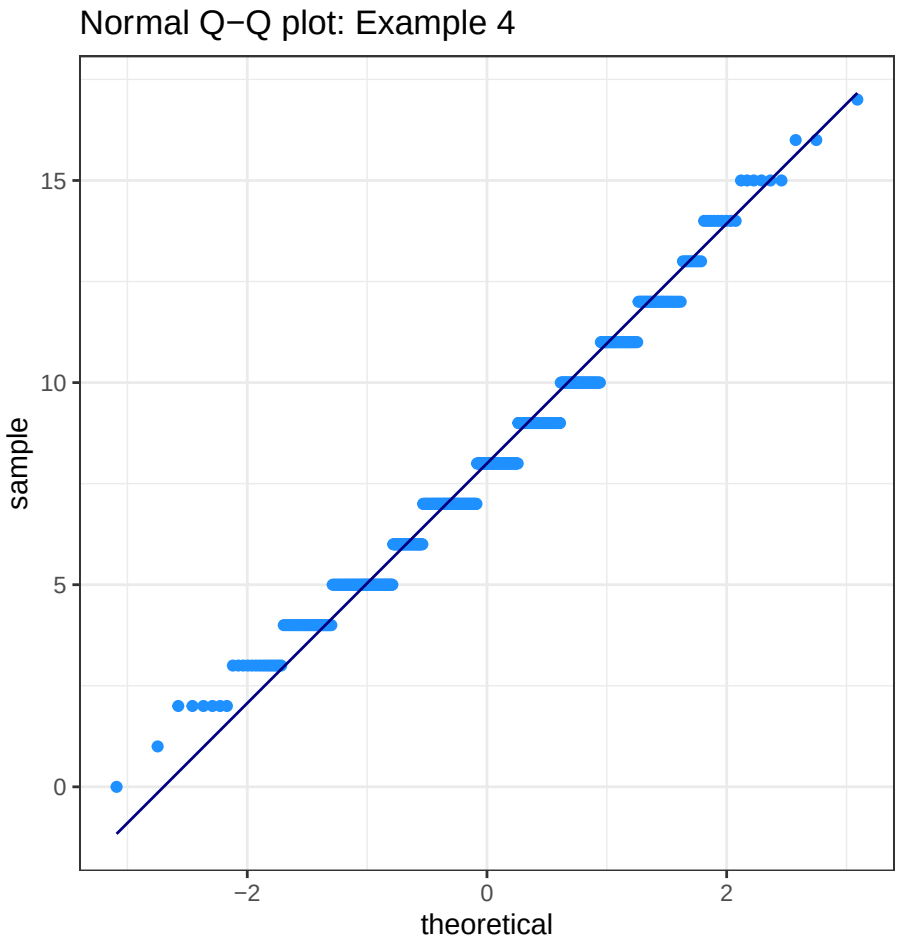


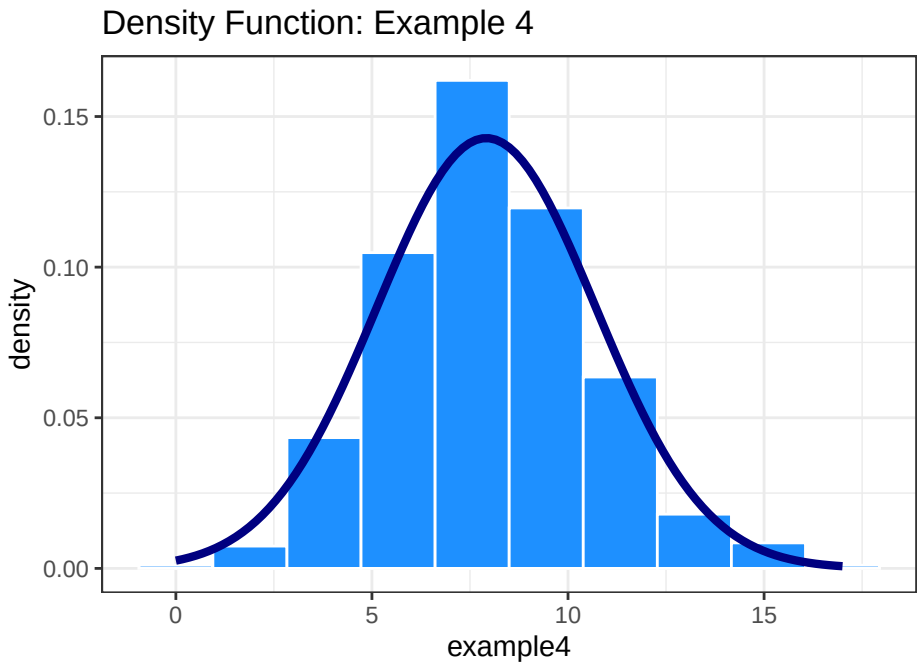


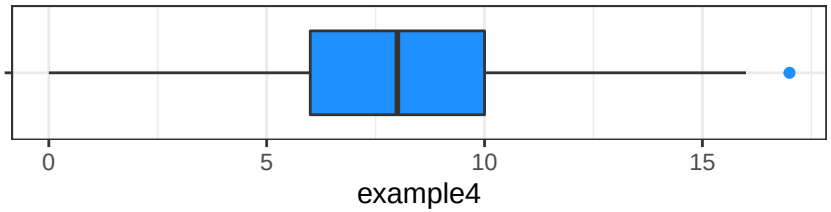


min	Q1	median	Q3	max	mean	sd	n	missing
0.5	1.9	2.6	3.8	11.6	3.1	1.7	500	0

Example 4: Discrete "Symmetric" Distribution

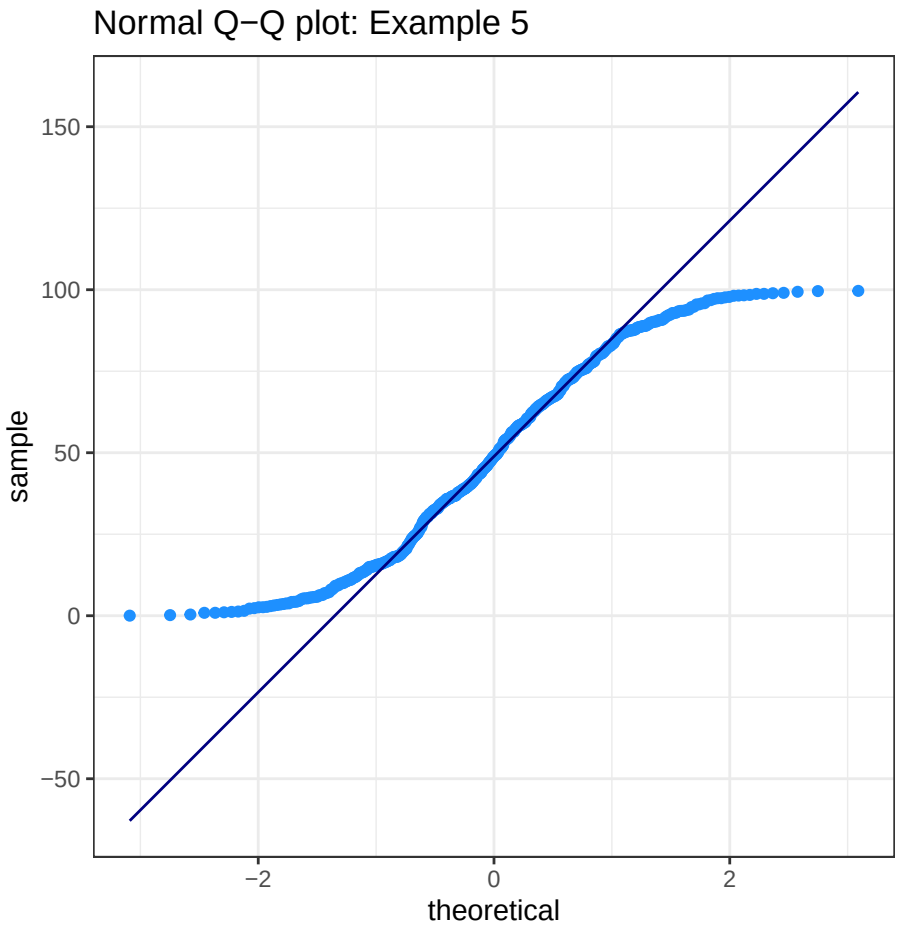


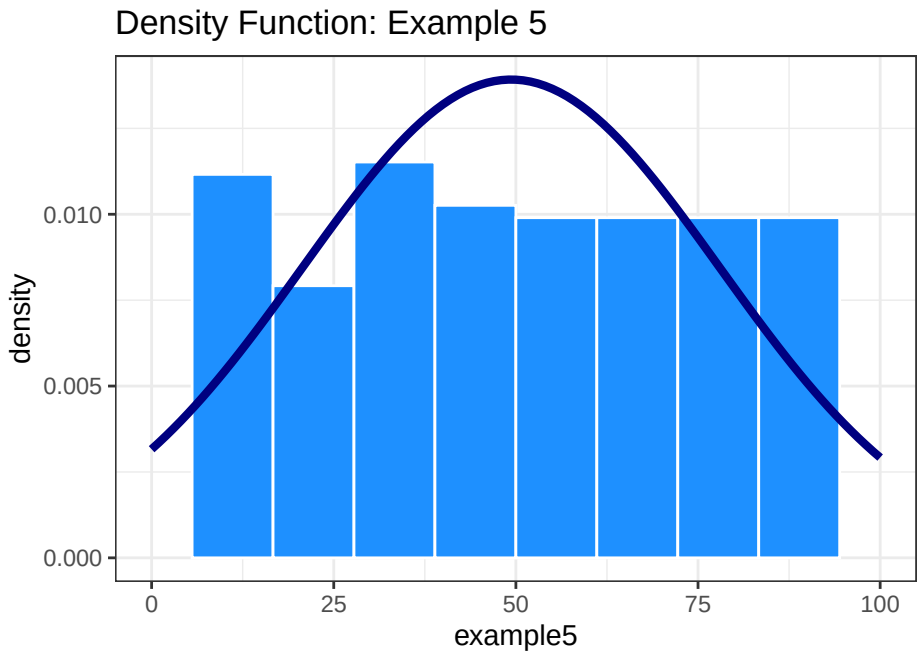


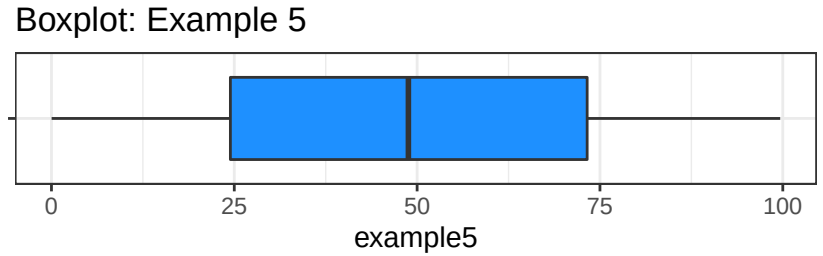


min	Q1	median	Q3	max	mean	sd	n	missing
0	6	8	10	17	7.9	2.8	500	0

Example 5: Data from a Uniform Distribution

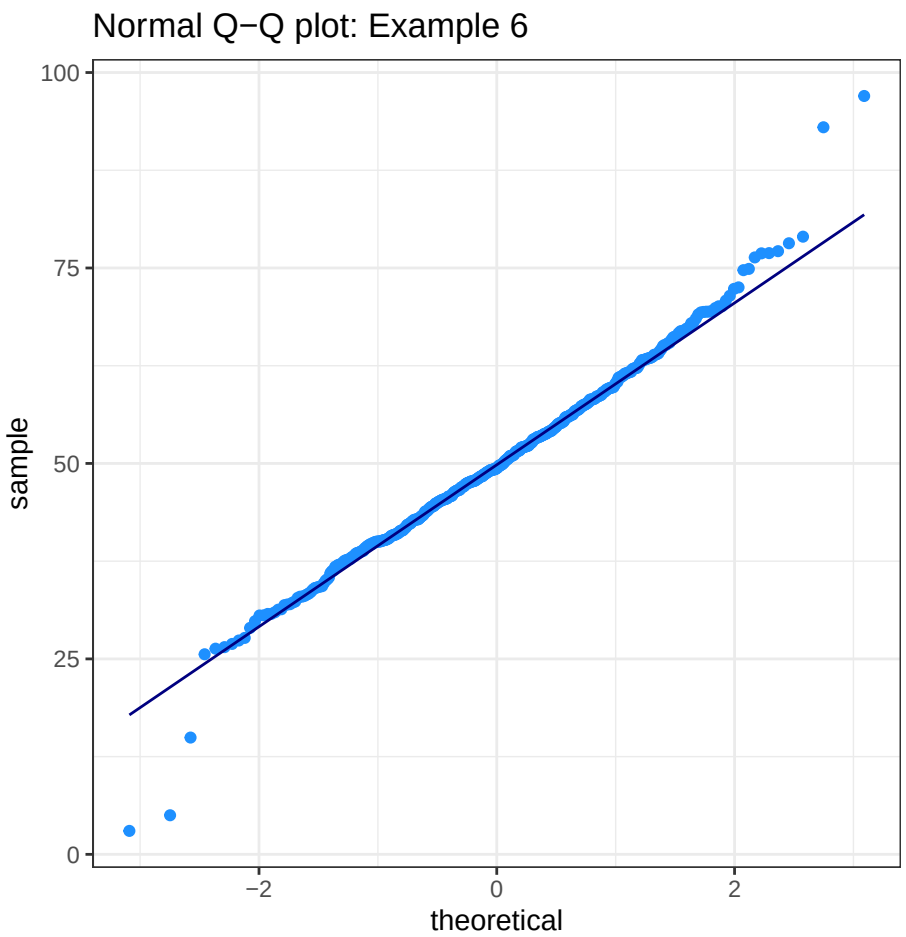


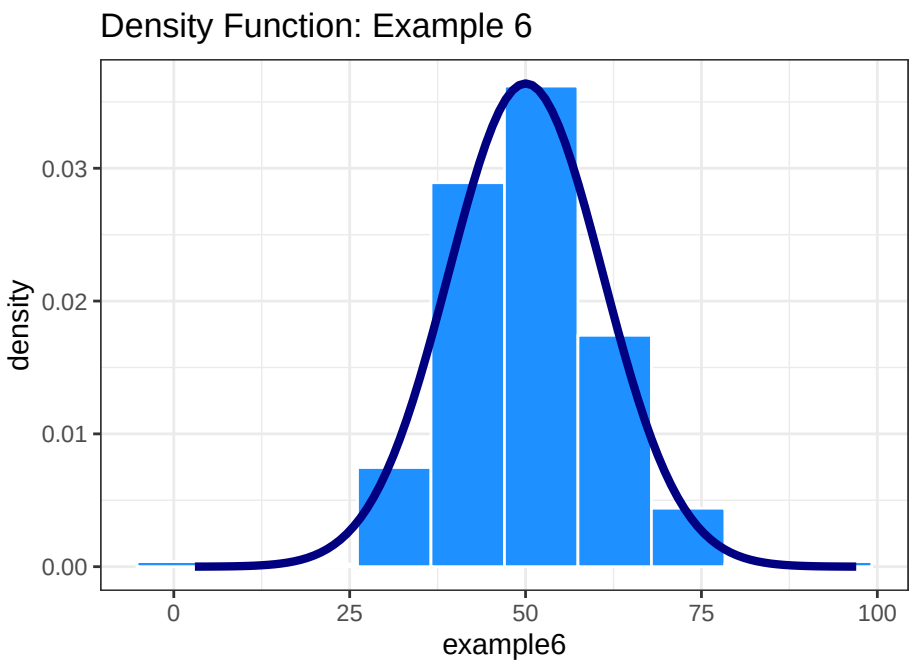


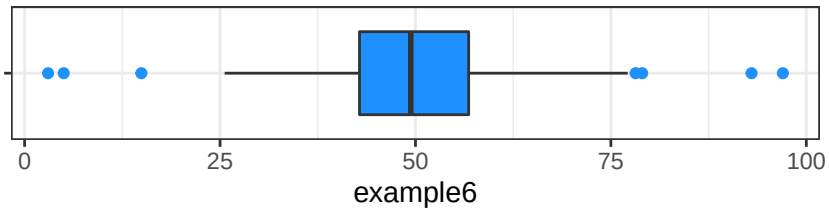


min	Q1	median	Q3	max	mean	sd	n	missing
0	24.5	48.8	73.2	99.6	49.3	28.7	500	0

Example 6: Symmetric data with outliers







min	Q1	median	Q3	max	mean	sd	n	missing
3	42.8	49.4	56.8	97	50	11	500	0

Returning to the blood pressure data

Today's Data and Agenda

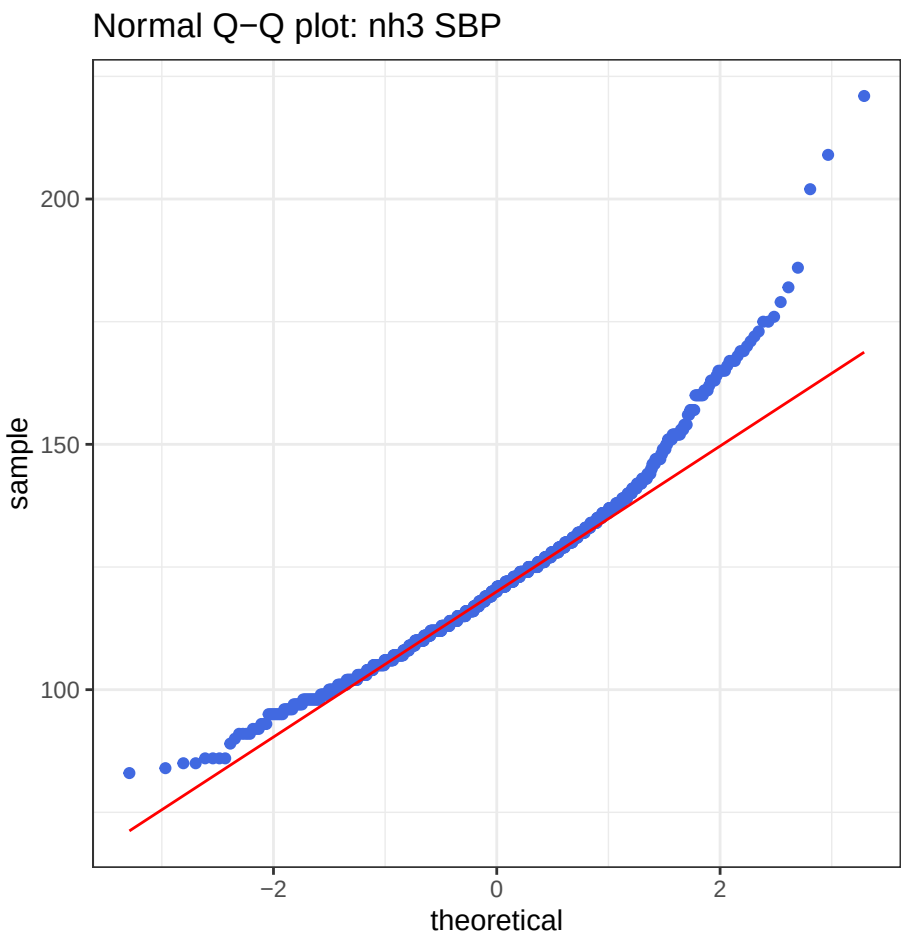
NHANES data from 2011-12: our `nh3` sample of 1000 adults.

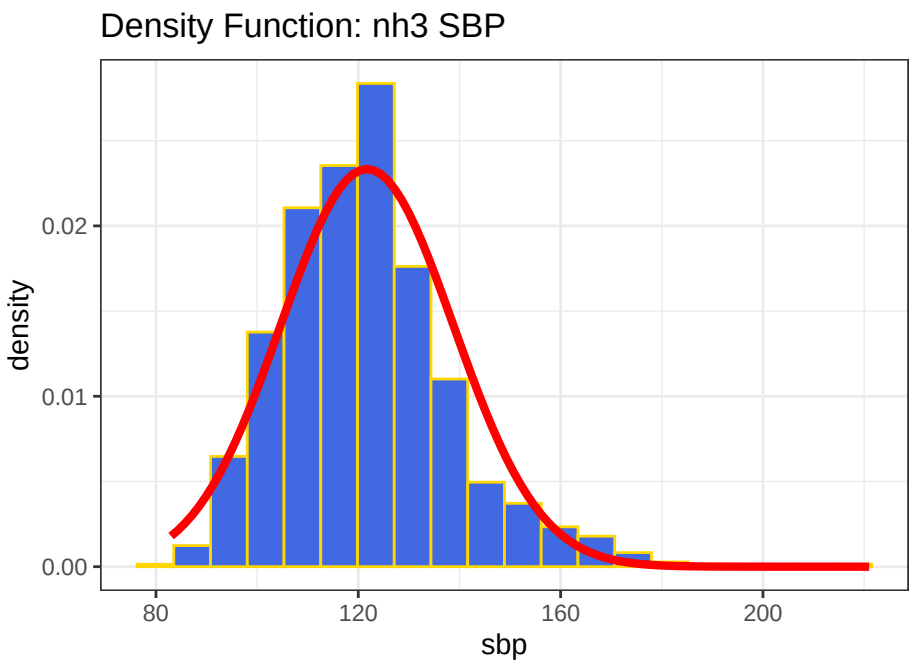
1. Are the systolic (SBP) and diastolic (DBP) blood pressure data in our sample well described by a Normal distribution?
2. How might we look at the association between SBP and DBP in our sample?

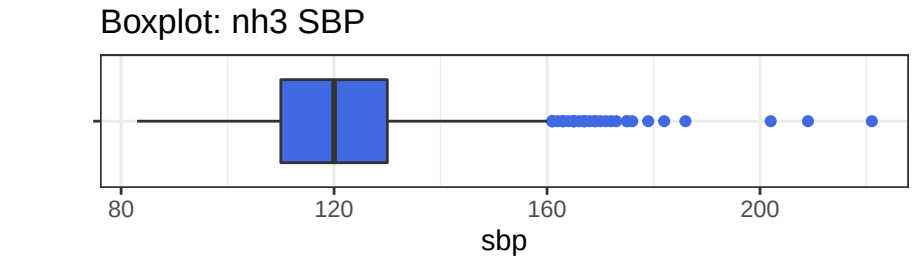
The nh3 data

```
set.seed(20200914)

nh3 <- NHANES %>%
    filter(SurveyYr == "2011_12") %>%
    select(ID, SurveyYr, Age, Height, Weight, BMI, Pulse,
           SleepHrsNight, BPSysAve, BPDiaAve, Gender,
           PhysActive, SleepTrouble, Smoke100,
           Race1, HealthGen, Depressed) %>%
    rename(SleepHours = SleepHrsNight, Sex = Gender,
           SBP = BPSysAve, DBP = BPDiaAve) %>%
    filter(Age > 20 & Age < 80) %>%
    drop_na() %>%
    distinct() %>%
    slice_sample(n = 1000) %>%
    clean_names()
```

Plots for the nh3 SBP data. How do we build this?







min	Q1	median	Q3	max	mean	sd	n	missing
83	110	120	130	221	121.7	17.1	1000	0

Code for sbp in nh3 (First of Three Plots)

```
p1 <- ggplot(nh3, aes(sample = sbp)) +
  geom_qq(col = "royalblue") +
  geom_qq_line(col = "red") +
  theme(aspect.ratio = 1) +
  labs(title = "Normal Q-Q plot: nh3 SBP")
```

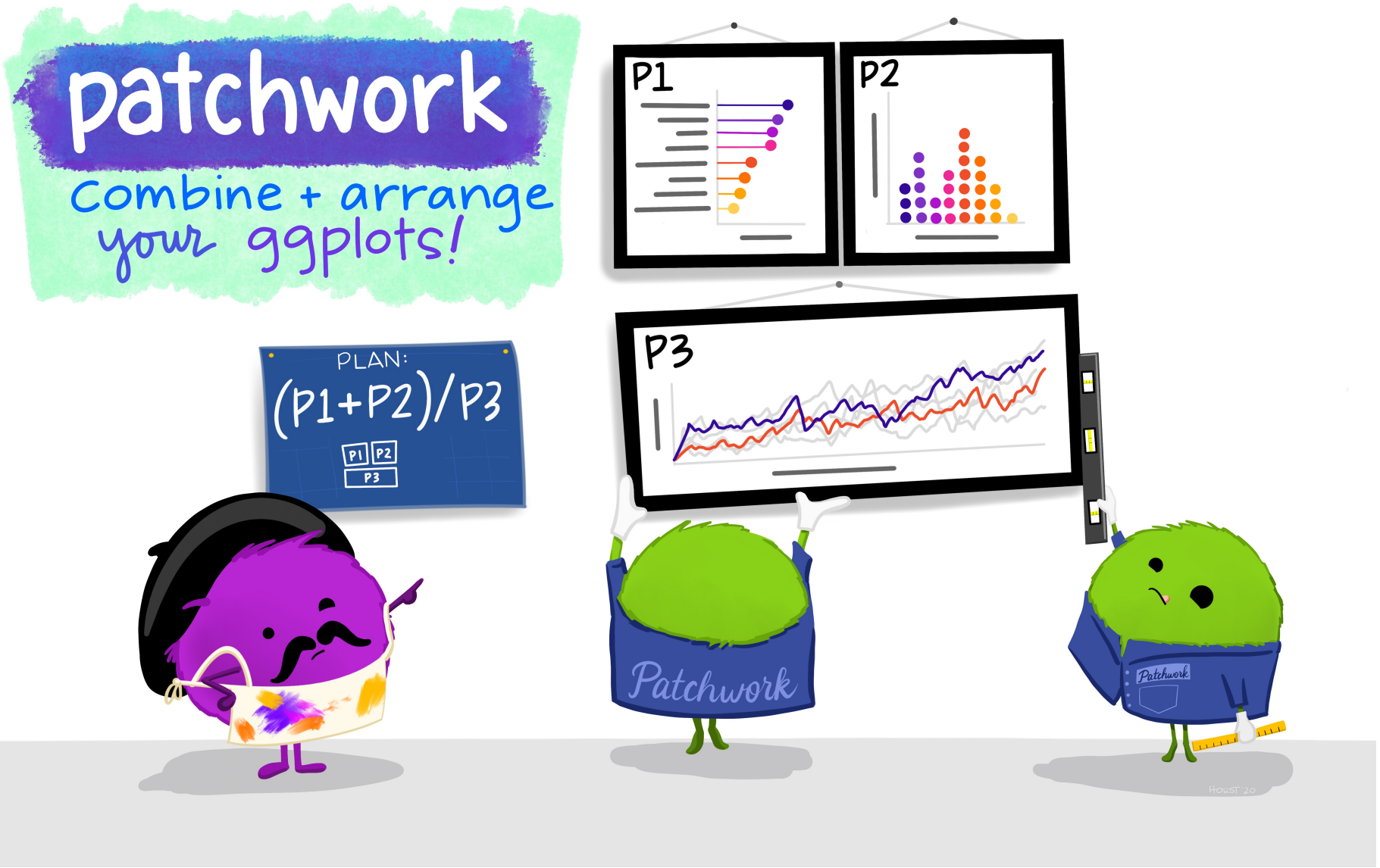
Code for sbp in nh3 (Second of Three Plots)

```
p2 <- ggplot(nh3, aes(x = sbp)) +
  geom_histogram(aes(y = stat(density)),
                 bins = 20,
                 fill = "royalblue", col = "gold") +
  stat_function(fun = dnorm,
                args = list(mean = mean(nh3$sbp),
                            sd = sd(nh3$sbp)),
                col = "red", lwd = 1.5) +
  labs(title = "Density Function: nh3 SBP")
```

Code for `sbp` in `nh3` (Third of Three Plots)

```
p3 <- ggplot(nh3, aes(x = sbp, y = "")) +
  geom_boxplot(fill = "royalblue",
               outlier.color = "royalblue") +
  labs(title = "Boxplot: nh3 SBP", y = "")
```

Putting the plots together...



Using patchwork

```
p1 + (p2 / p3 + plot_layout(heights = c(4,1)))
```

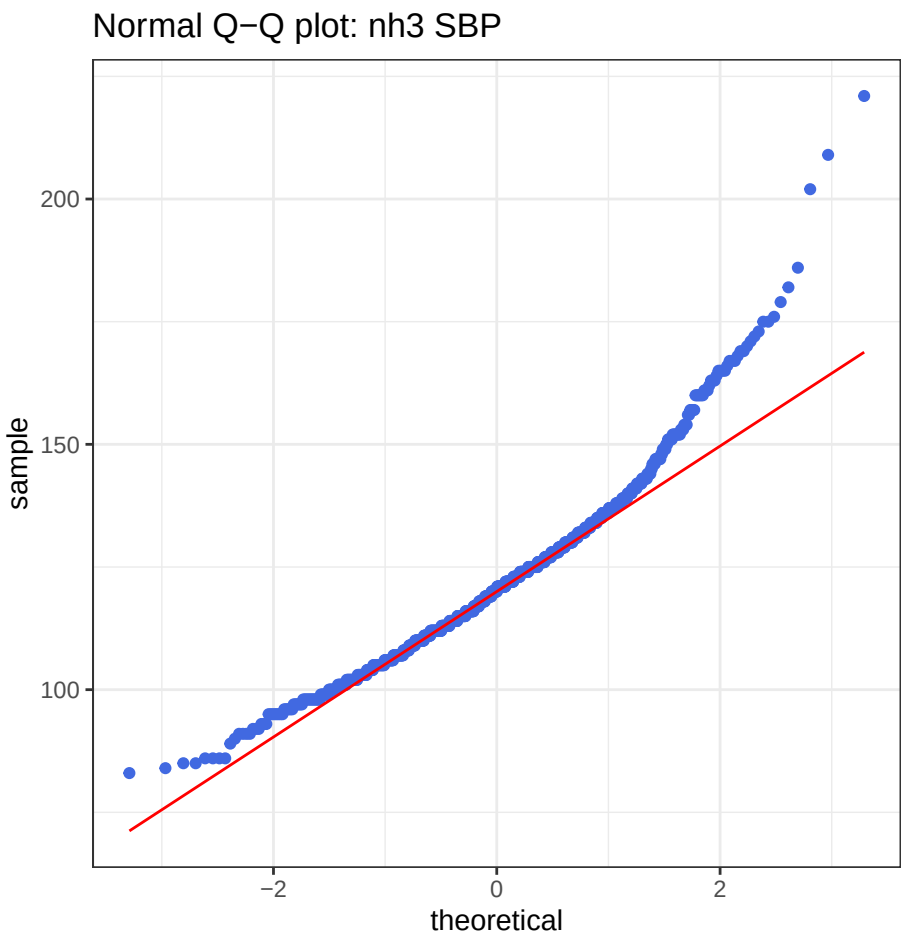
Learn more about **patchwork** at https://patchwork.data-imaginist.com/. In particular, the package can also be used to add annotations to your plots.

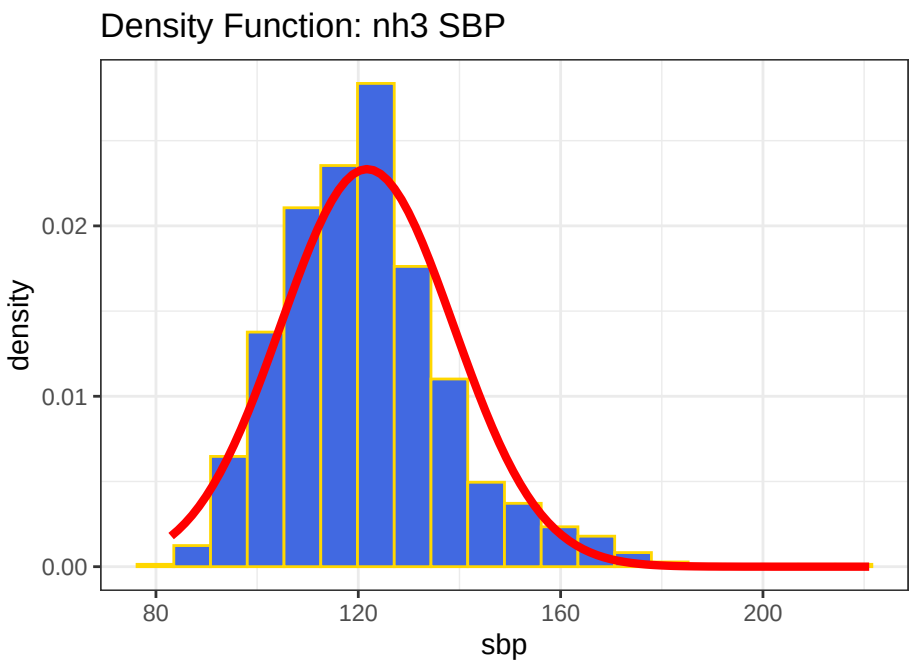
Here, outside of patchwork, I added...

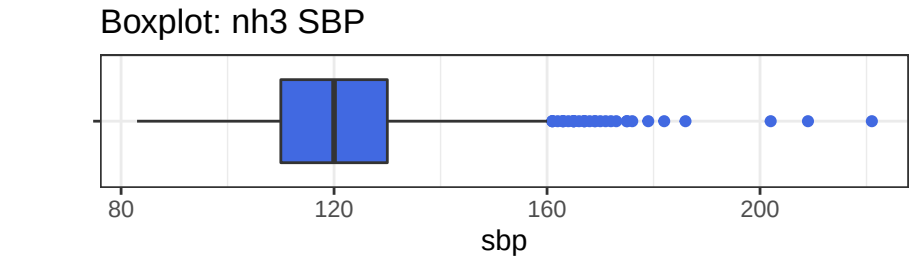
```
mosaic::favstats(~ sbp, data = nh3) %>% kable(digits = 1)
```

min	Q1	median	Q3	max	mean	sd	n	missing
83	110	120	130	221	121.7	17.1	1000	0

Result: 1000 observed Systolic BP values







min	Q1	median	Q3	max	mean	sd	n	missing
83	110	120	130	221	121.7	17.1	1000	0

More Extensive Numerical Summaries?

We could try

```
nh3 %>% Hmisc::describe(sbp)
```

- but that will throw an error message, specifically `Error in describe.data.frame(., sbp) : object 'sbp' not found`. What is wrong? How can we fix that?

More Extensive Numerical Summaries?

We could try

```
nh3 %>% Hmisc::describe(sbp)
```

- but that will throw an error message, specifically `Error in describe.data.frame(., sbp) : object 'sbp' not found`. What is wrong? How can we fix that?
- We could drop the pipe and use `$` notation, so `Hmisc::describe(nh3$sbp)`

More Extensive Numerical Summaries?

We could try

```
nh3 %>% Hmisc::describe(sbp)
```

- but that will throw an error message, specifically `Error in describe.data.frame(., sbp) : object 'sbp' not found`. What is wrong? How can we fix that?
- We could drop the pipe and use `$` notation, so `Hmisc::describe(nh3$sbp)`
- Another option is to change the pipe (to the `%$%` pipe available in the `magrittr` package): `nh3 %$% Hmisc::describe(sbp)`

More Extensive Numerical Summaries?

We could try

```
nh3 %>% Hmisc::describe(sbp)
```

- but that will throw an error message, specifically `Error in describe.data.frame(., sbp) : object 'sbp' not found`. What is wrong? How can we fix that?
- We could drop the pipe and use `$` notation, so `Hmisc::describe(nh3$sbp)`
- Another option is to change the pipe (to the `%$%` pipe available in the `magrittr` package): `nh3 %$% Hmisc::describe(sbp)`
- We use the `%>%` pipe most of the time within the tidyverse, but will need `%$%` sometimes for functions (like those in `Hmisc`) that are less tidy-friendly.

What do these summaries tell us?

```
nh3 %$% Hmisc::describe(sbp)
```

```
sbp
       n  missing distinct     Info     Mean      Gmd
    1000        0       93        1    121.7    18.52
     .05      .10      .25      .50      .75      .90
      98      102      110      120      130      142
     .95
     152

lowest :  83  84  85  86  89, highest: 182 186 202 209 221
```

- `Gmd` = Gini's mean difference (a robust measure of spread) = mean absolute difference between any pairs of observations. Larger `Gmd` indicates more spread.
- `Info` = a measure of relative information describing how "continuous" the data are. Higher `Info` indicates fewer ties (more distinct values.)

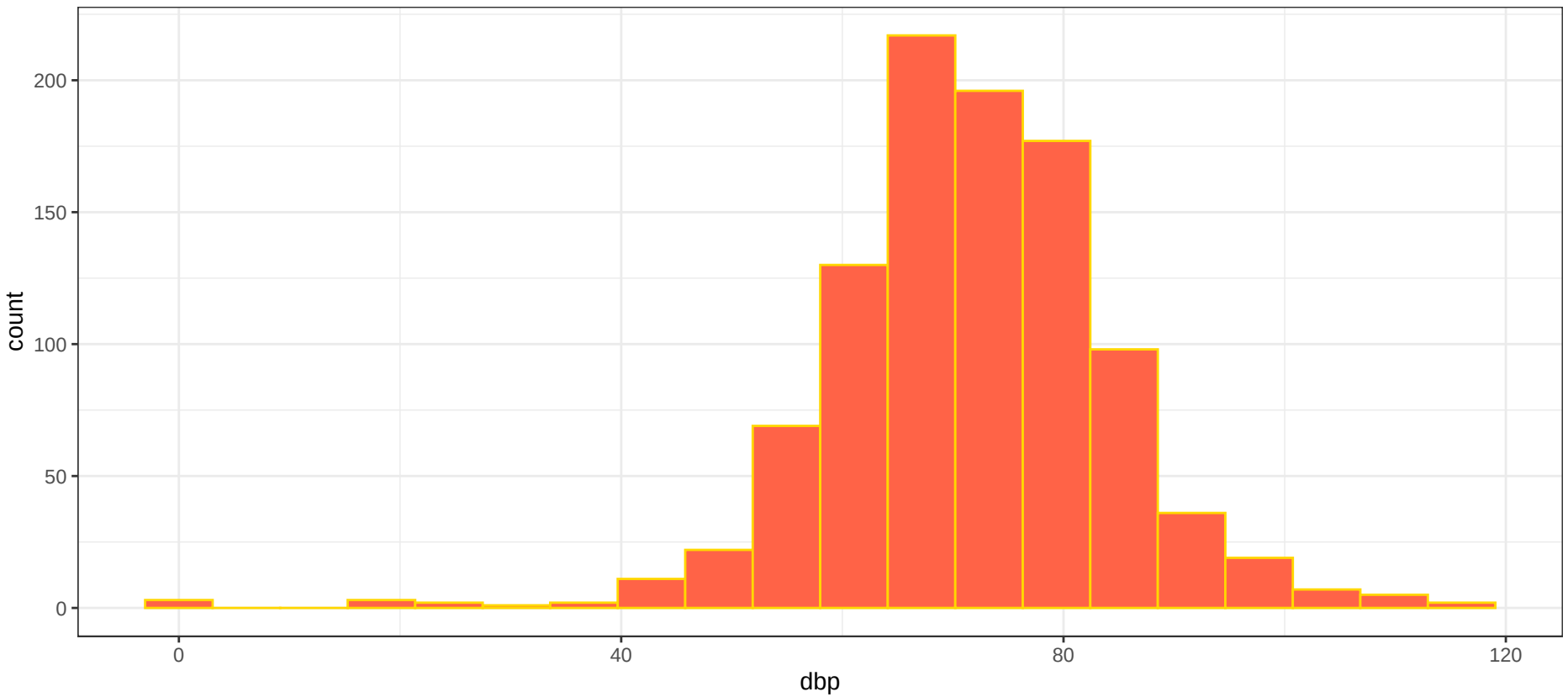
What Summaries to Report

Focus on the shape, center and spread of a distribution. From Bock, Velleman and DeVeaux. . .

- If the data are **skewed**, report the median and IQR (or the three middle quantiles). You may want to include the mean and standard deviation, but you should point out why the mean and median differ. The fact that the mean and median do not agree is a sign that the distribution may be skewed. A histogram will help you make that point.
- If the data are **symmetric**, report the mean and standard deviation, and possibly the median and IQR as well.
- If there are **clear outliers** and you are reporting the mean and standard deviation, report them with the outliers present and with the outliers removed. The differences may be revealing. The median and IQR are not likely to be seriously affected by outliers.

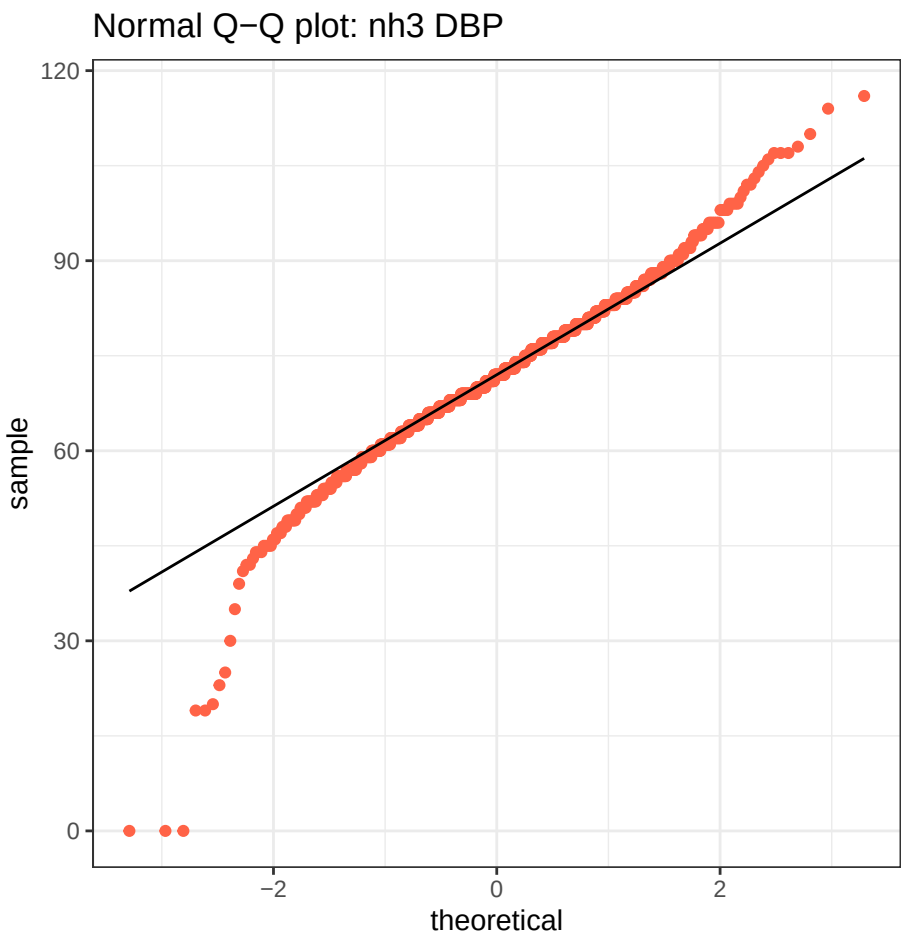
OK, what about Diastolic Blood Pressure?

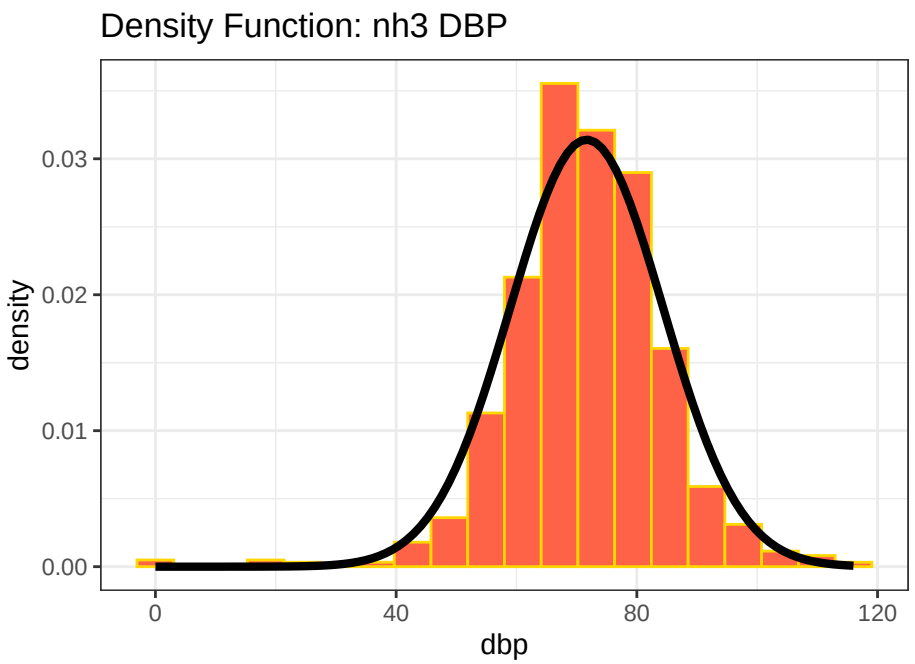
```
ggplot(data = nh3, aes(x = dbp)) +
  geom_histogram(bins = 20, fill = "tomato", col = "gold")
```

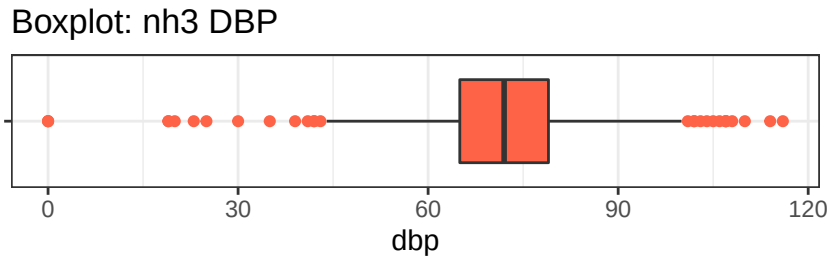


- We'll generate the exploratory summaries we've been using. . .

DBP in nh3: Center/Spread/Outliers/Shape?







min	Q1	median	Q3	max	mean	sd	n	missing
0	65	72	79	116	71.7	12.7	1000	0

Hmisc::describe for dbp?

```
nh3 %$% Hmisc::describe(dbp)
```

```
dbp
       n  missing distinct     Info     Mean      Gmd
    1000        0       78    0.999    71.66    13.58
     .05      .10      .25      .50      .75      .90
      52       57       65       72       79       86
     .95
      91

lowest :   0  19  20  23  25, highest: 107 108 110 114 116
```

What is a plausible diastolic blood pressure?

Stem-and-Leaf of dbp values?

```
stem(nh3$dbp)
```

```
The decimal point is 1 digit(s) to the right of the |

 0 | 000
 1 | 99
 2 | 035
 3 | 059
 4 | 1223444555566777888999999
 5 | 000111112222222233333334444444455555556666666666666777
 6 | 000000000000000011111111111111111111222222222222222
 7 | 00000000000000000000000000001111111111111111111111
 8 | 000000000000000000000000000011111111111111112222
 9 | 00000000111111222222334444445555666666888899999
10 | 012234567778
11 | 046
```

Who are those people with tiny dbp values?

```
nh3 %>%
  filter(dbp < 40) %>%
  select(id, sbp, dbp)
```

```
# A tibble: 11 x 3
      id   sbp   dbp
   <int> <int> <int>
 1 71598    86    30
 2 68528   133    39
 3 64298   135     0
 4 64616   111    25
 5 65298   126    35
 6 62649   122    23
 7 70664   152     0
 8 69237   120     0
 9 68561   119    19
10 68908   129    20
```

Let's reset.

```
nh3_new <- nh3 %>%
  filter(dbp > 39)

nrow(nh3)
```

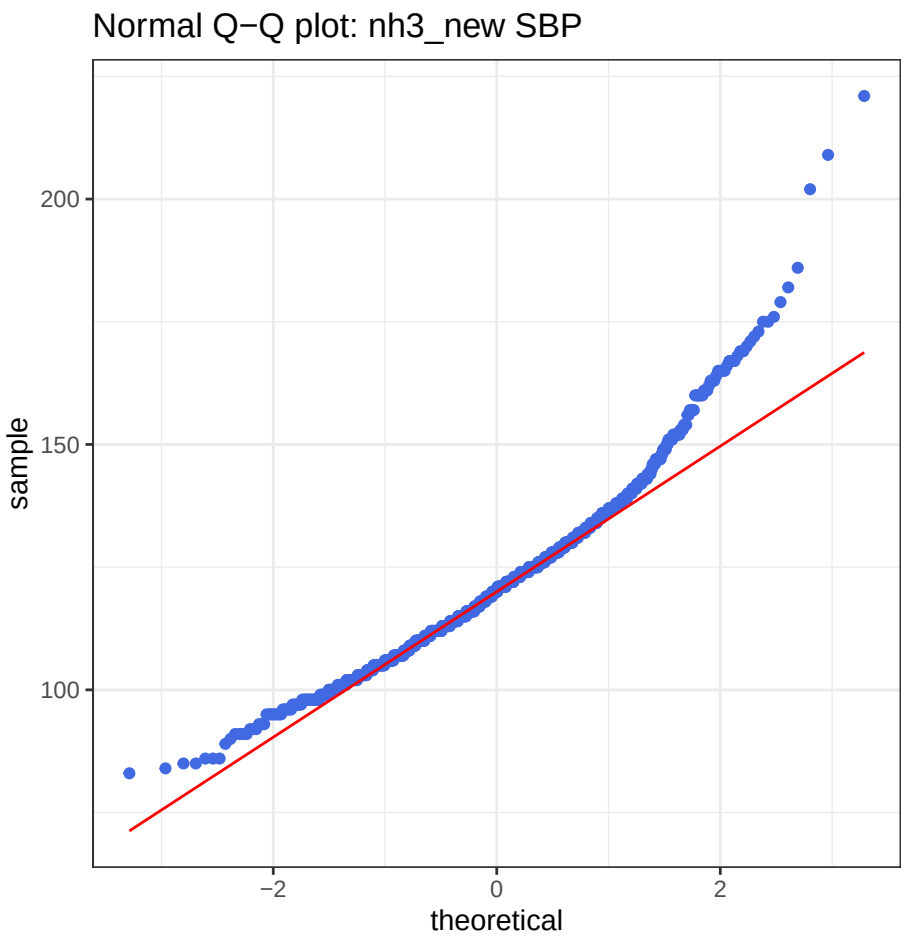
```
[1] 1000
```

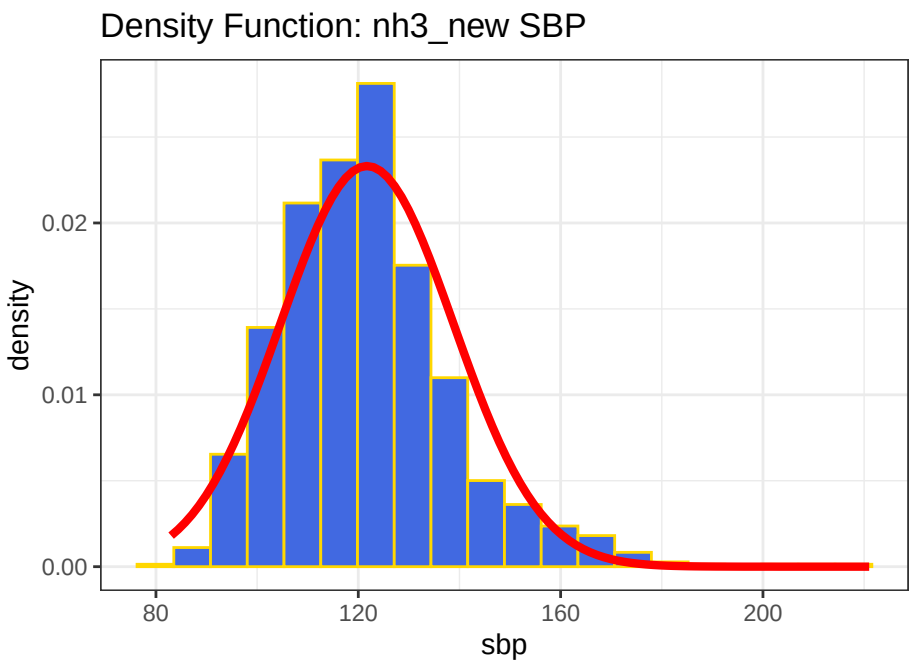
```
nrow(nh3_new)
```

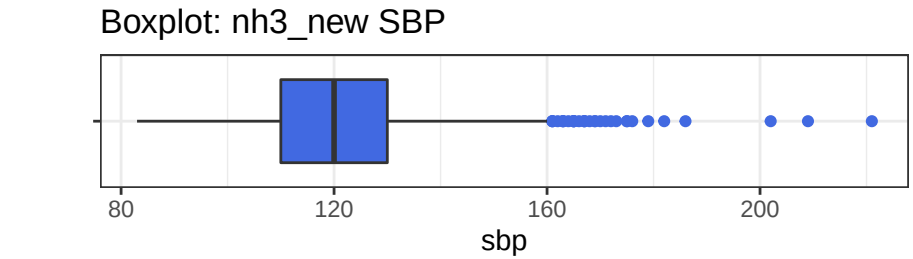
```
[1] 989
```

We'll work with `nh3_new` going forward.

nh3_new: Systolic Blood Pressure

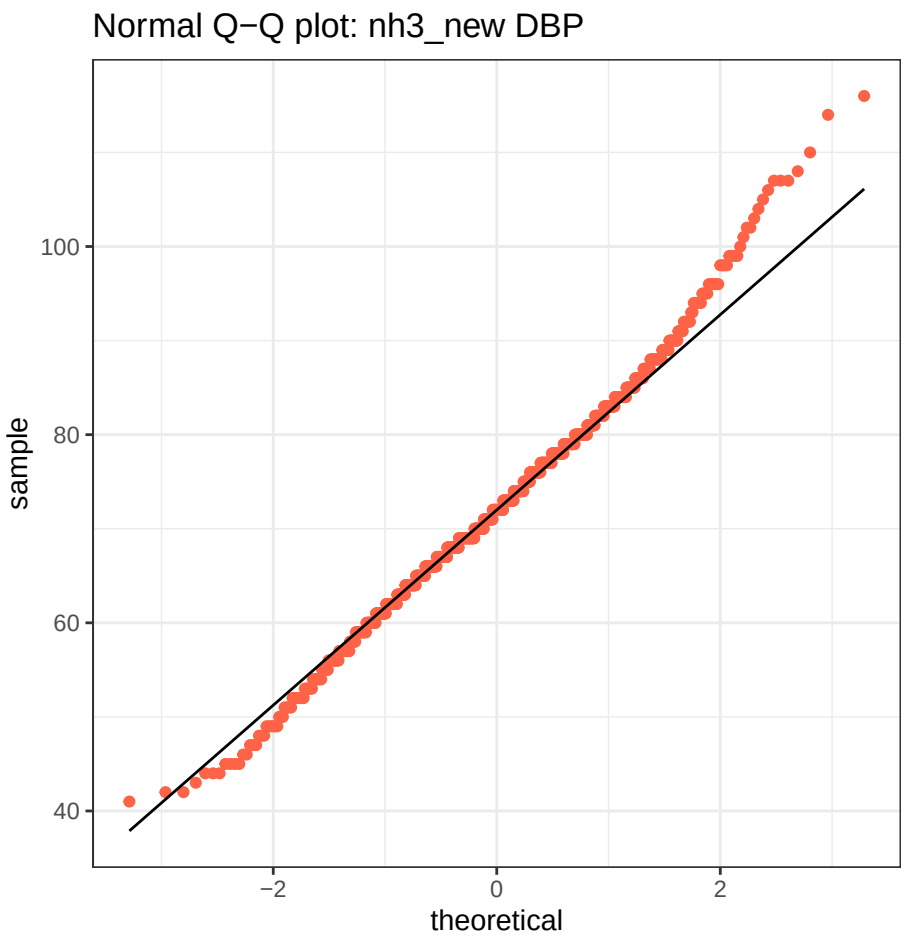


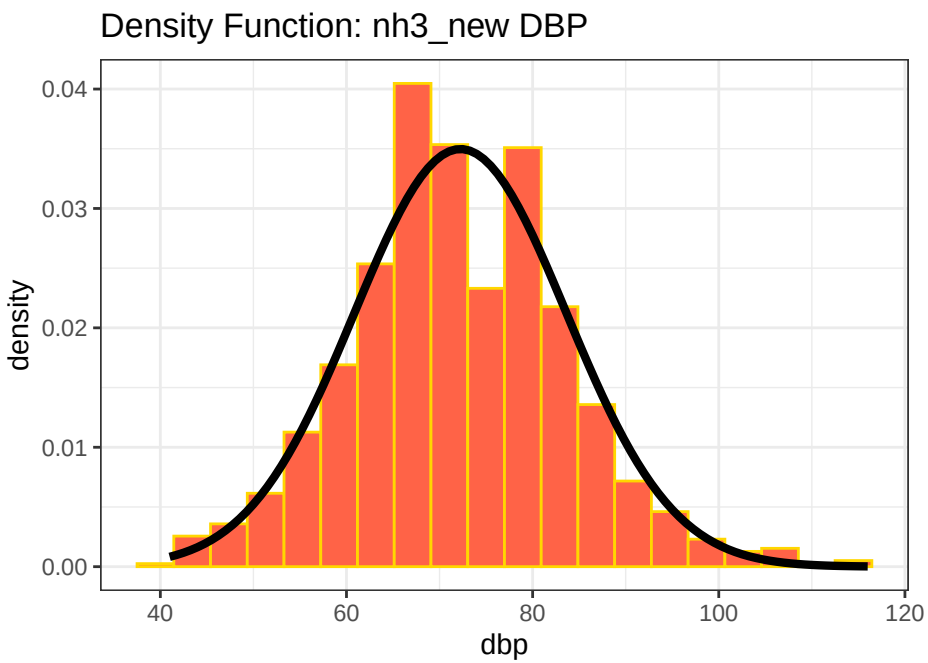


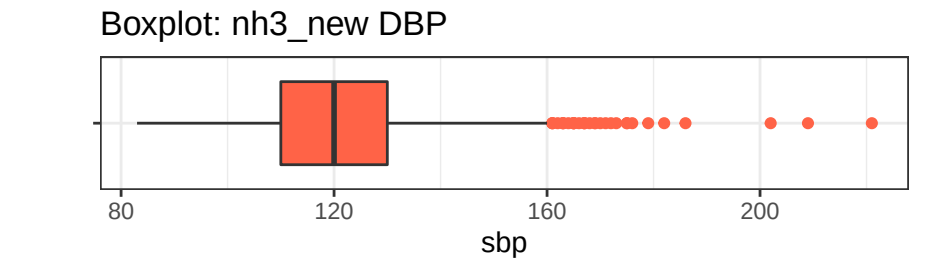


min	Q1	median	Q3	max	mean	sd	n	missing
83	110	120	130	221	121.7	17.1	989	0

nh3_new: Diastolic Blood Pressure







min	Q1	median	Q3	max	mean	sd	n	missing
41	65	72	79	116	72.2	11.4	989	0

Summing Up: Does a Normal Model fit well?

If a Normal model fits our data well, then we should see the following graphical indications:

1. A histogram that is symmetric and bell-shaped.
2. A boxplot where the box is symmetric around the median, as are the whiskers, without a serious outlier problem.
3. A normal Q-Q plot that essentially falls on a straight line.

As for numerical summaries, we'd like to see

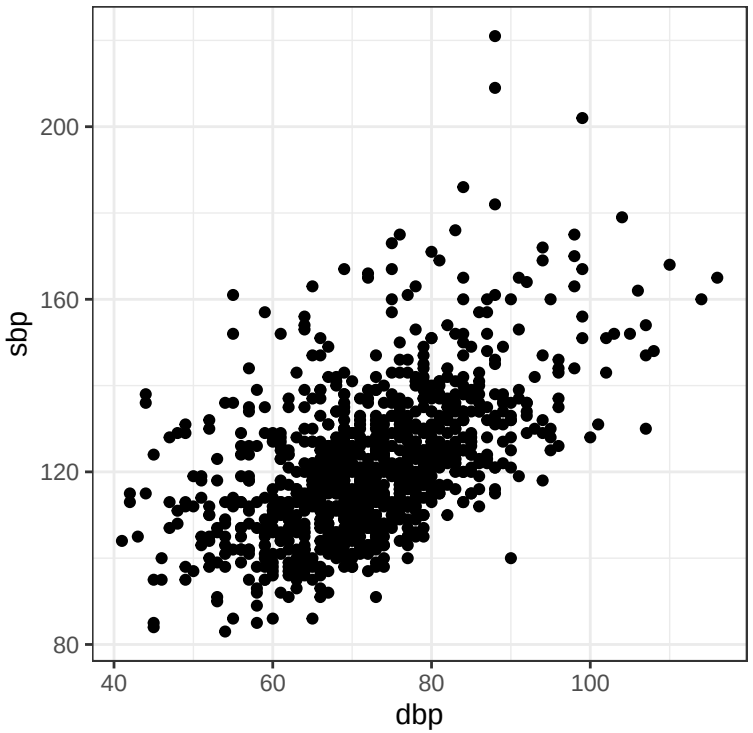
4. The mean and median within 0.2 standard deviation of each other.
5. No real evidence of too many outlier candidates (more than 5% starts to get us concerned about a Normal model)
6. No real evidence of individual outliers outside the reasonable range for the size of our data (we might expect about 3 observations in 1000 to fall more than 3 standard deviations away from the mean.)

How can we describe the relationship between SBP and DBP?

Scatterplot to study the SBP-DBP association

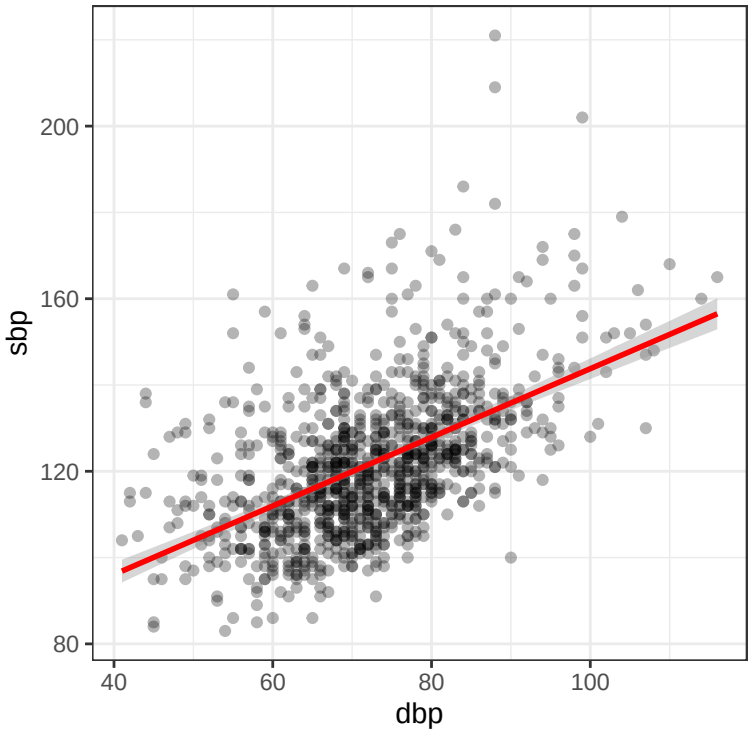
One point for each of the 989 subjects in the `nh3_new` data set. . .

```
ggplot(nh3_new, aes(x = dbp, y = sbp)) +
  geom_point() +
  theme(aspect.ratio = 1) # make plot square
```

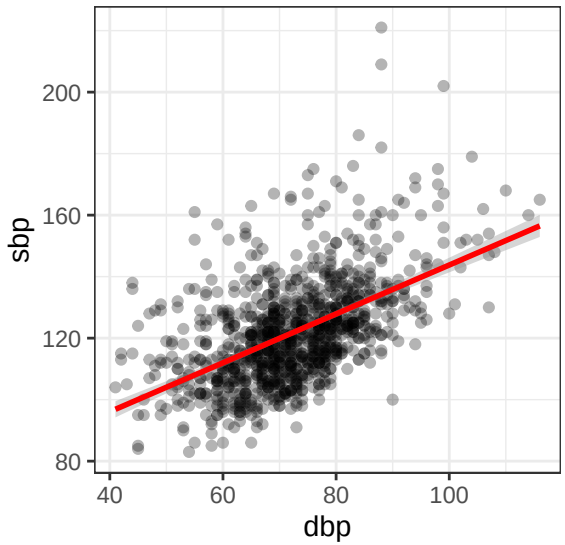


Add a fitted regression line?

```
ggplot(nh3_new, aes(x = dbp, y = sbp)) +
  geom_point(alpha = 0.3) + # add some transparency
  theme(aspect.ratio = 1) +
  geom_smooth(method = "lm", formula = y ~ x,
              col = "red", se = TRUE)
```



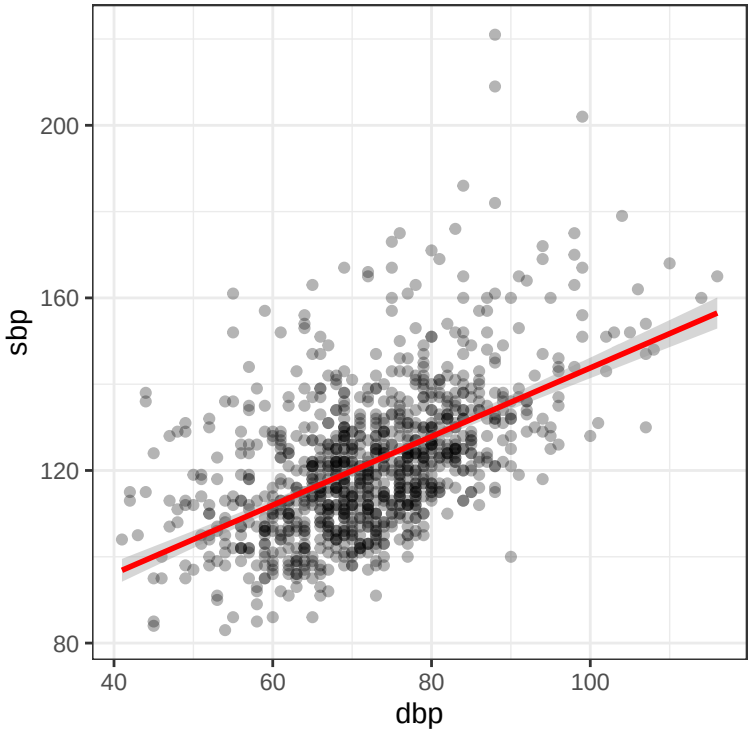
What are we looking for in this plot?



Is the association. . .

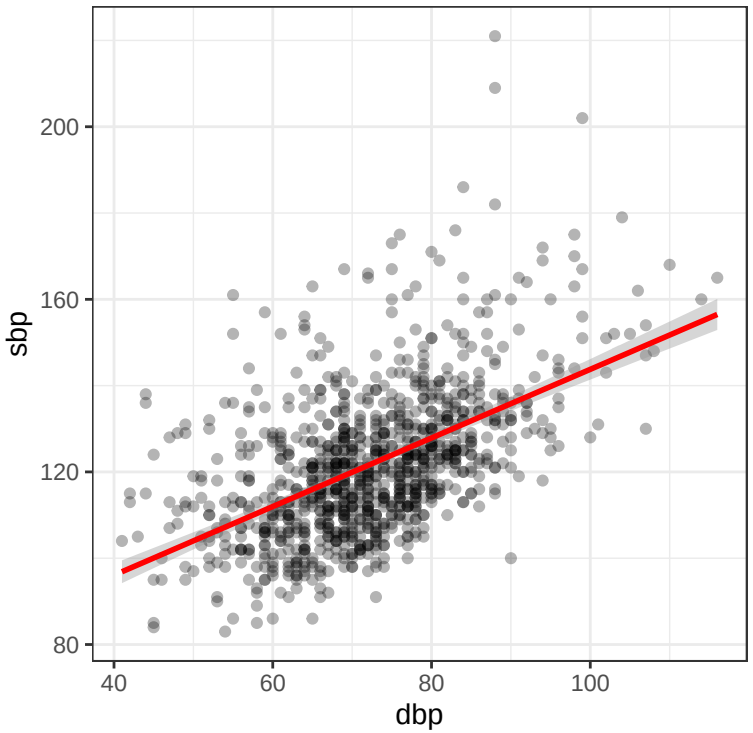
1. **Linear or Non-Linear?** (is there a curve here?)
2. **Direction?** (as X increases, what happens to Y?)
3. **Outliers?** (far away on X, or Y, or the combination?)
4. **Strength?** (points closely clustered together around a line?)

What might we conclude here?



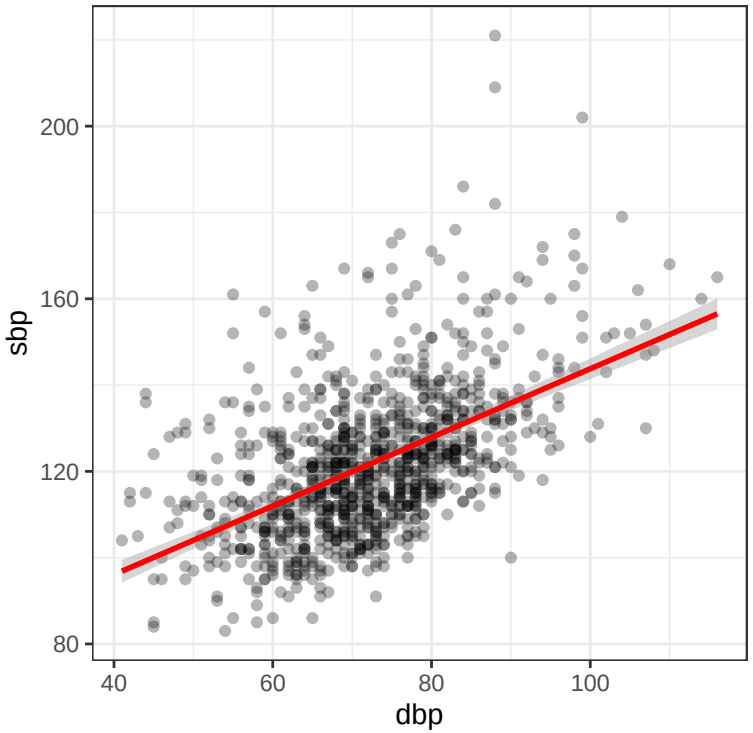
1. **Linear?**: Most of the points cluster around the straight line.

What might we conclude here?



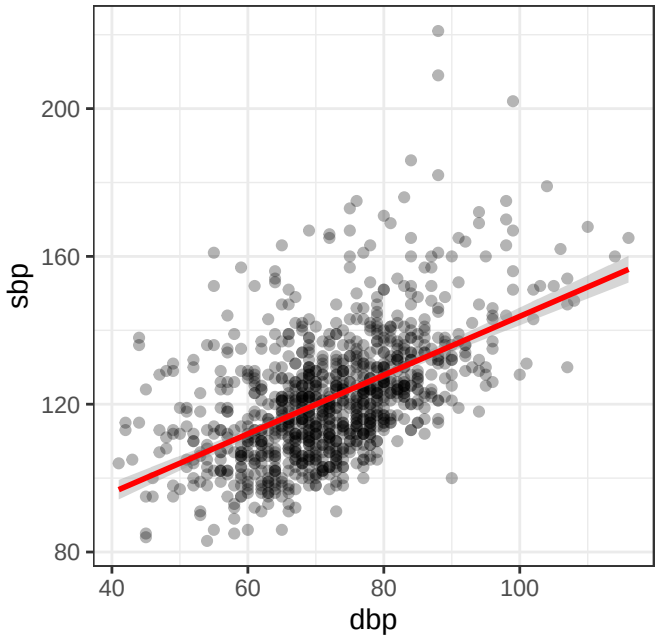
1. **Linear?**: Most of the points cluster around the straight line.
2. **Direction?**: As dbp increases, so does sbp, generally.

What might we conclude here?



1. **Linear?**: Most of the points cluster around the straight line.
2. **Direction?**: As dbp increases, so does sbp, generally.
3. **Outliers?**: A few (out of 1000) worth another look, probably.

What might we conclude here?



4. **Strength?**: The association does not seem very strong.

- The range of `sbp` values associated with any particular `dbp` value is still pretty wide.
- If we know someone's `dbp`, that should help us make better predictions of their `sbp`, but maybe only a little better than if we didn't know `dbp`.

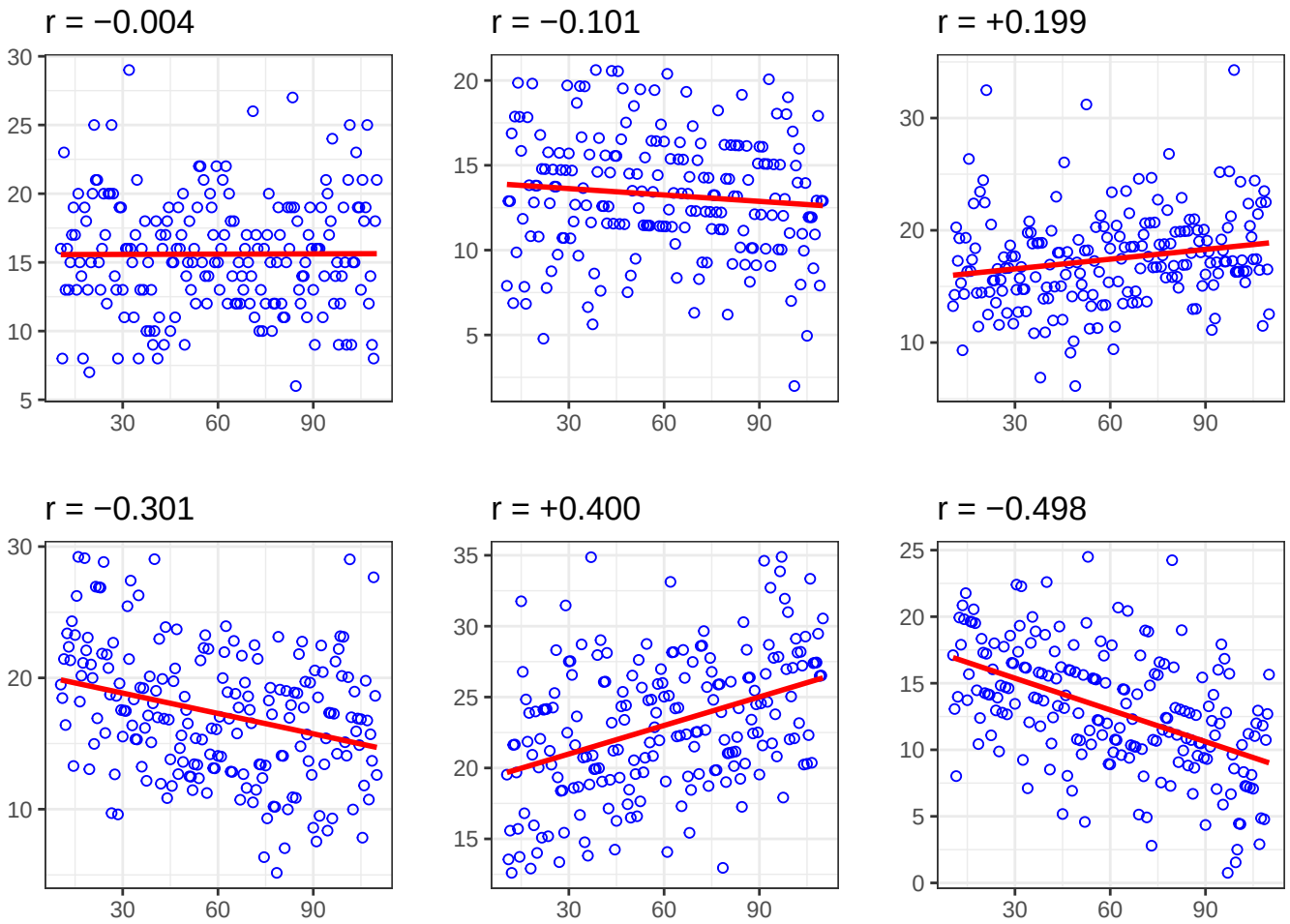
What do you think the correlation of `sbp` and `dbp` might be?

Summarizing Strength with the Pearson Correlation

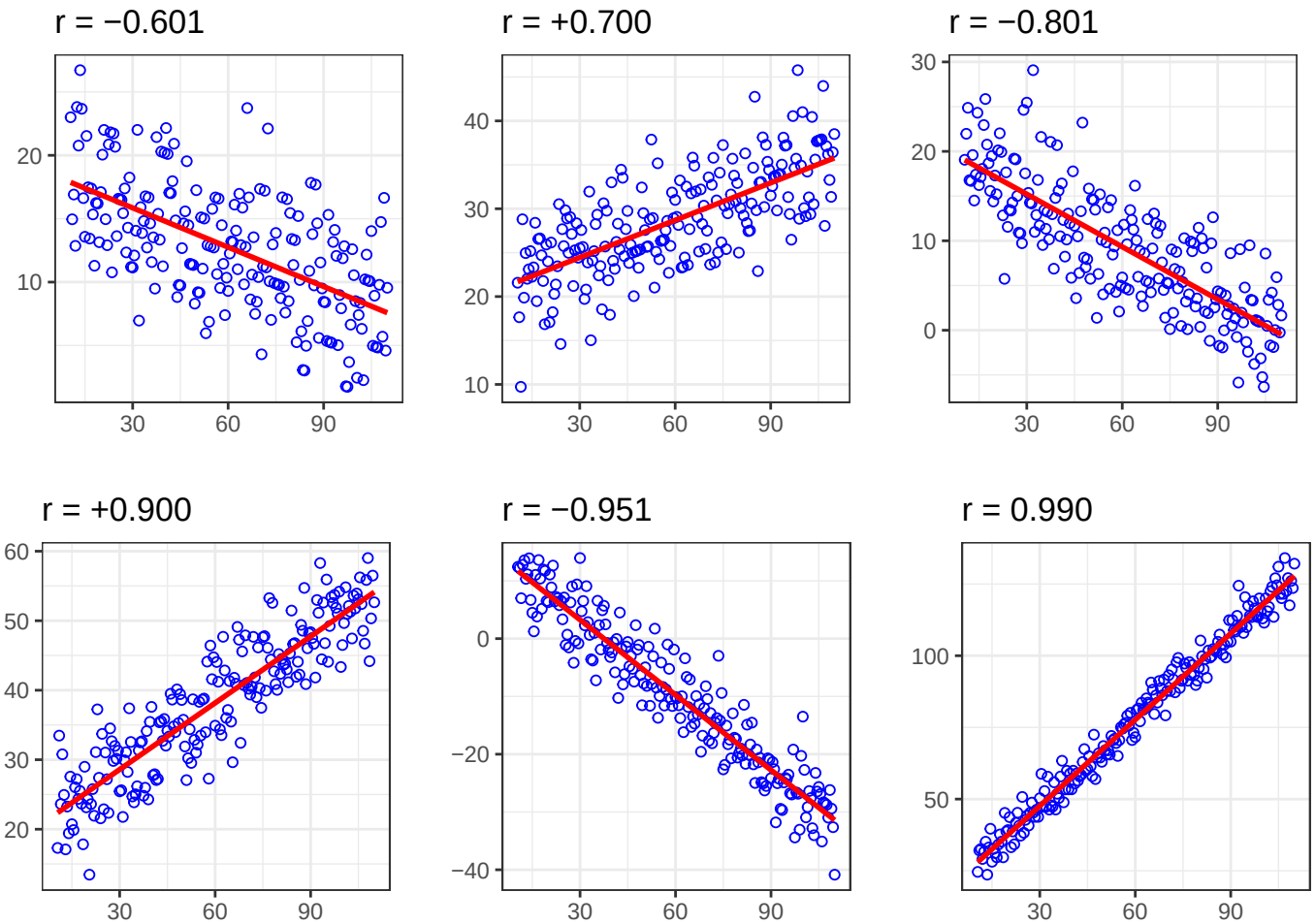
The Pearson correlation (abbreviated r) ranges from -1 to $+1$.

- The closer the absolute value of the correlation is to 1, the stronger a linear fit will be to the data, (in a limited sense).
- A strong positive correlation (near $+1$) will indicate a strong model with a positive slope.
- A strong negative correlation (near -1) will indicate a strong linear model with a negative slope.
- A weak correlation (near 0) will indicate a poor fit for a linear model, although a non-linear model may still fit the data quite well.

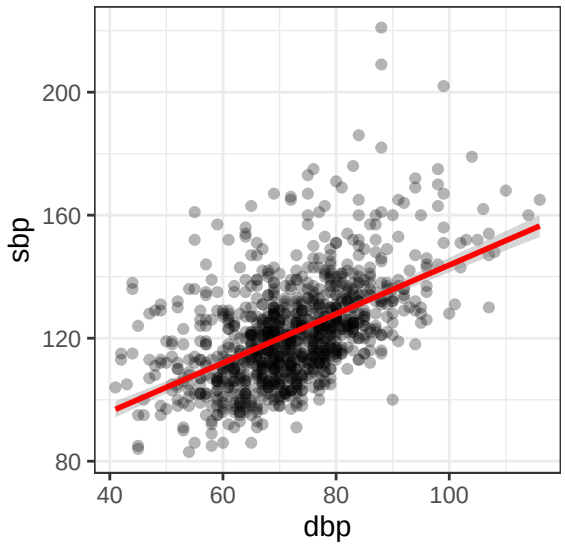
Gaining Some Insight into Correlation



Some Stronger Correlations



Correlation in our sbp-dbp scatterplot?




```
nh3_new %$% cor(sbp, dbp)
```

```
[1] 0.5299471
```

What does a correlation of +0.53 imply about a linear fit to the data?

What line is being fit?

Least Squares Regression Line (a linear model) to predict sbp using dbp

```
m1 <- lm(sbp ~ dbp, data = nh3_new)
m1
```

```
Call:
lm(formula = sbp ~ dbp, data = nh3_new)

Coefficients:
(Intercept)          dbp
     64.270        0.795
```

Model m1 is **sbp = 64.270 + 0.795 dbp**.

What does the slope mean?

$$\underset{\text{(tons)}}{\text{weight}} = 2.4 + 0.3(\text{height}) + \ldots$$

if all other variables constant, we expect a 1 foot taller dragon to weigh 0.3 tons more, on average.

Linear Model `m1`: sbp = 64.27 + 0.795 dbp

64.27 is the intercept = predicted value of `sbp` when `dbp` = 0.

0.795 is the slope = predicted change in `sbp` per 1 unit change in `dbp`

- What are the units?
- What does the fact that this estimated slope is positive mean?
- What would the line look like if the slope was negative?
- What if the slope was zero?

Summarizing the Fit

The summary function when applied to a linear model (lm) produces a lot of output that is not organized in a way that we can plot/manipulate it well.

Here's the start of what it looks like. . . (complete snapshot on next slide)

```
summary(m1)
```

```
Call:
lm(formula = sbp ~ dbp, data = nh3_new)

Residuals:
    Min      1Q  Median      3Q     Max
-35.824  -9.792  -2.103   6.947  86.766

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.2698     2.9617   21.70   <2e-16 ***
```

summary(m1) in its entirety

```
> summary(m1)

Call:
lm(formula = sbp ~ dbp, data = nh3_new)

Residuals:
    Min      1Q  Median      3Q     Max
-35.824  -9.792  -2.103   6.947  86.766

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.2698     2.9617   21.70   <2e-16 ***
dbp           0.7950     0.0405   19.63   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.52 on 987 degrees of freedom
Multiple R-squared:  0.2808,    Adjusted R-squared:  0.2801
F-statistic: 385.4 on 1 and 987 DF,  p-value: < 2.2e-16
```

Why I like `tidy()` and other `broom` functions

https://github.com/allisonhorst/stats-illustrations

Does R like this linear model?

```
tidy(m1) %>% kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	64.27	2.96	21.70	0
dbp	0.80	0.04	19.63	0

Yes. Wow. It **really** does. Look at those *p* values!

How much of the variation in sbp does m1 capture?

The glance function can help us (again from broom.)

```
glance(m1) %>% select(r.squared, p.value, sigma) %>% kable()
```

r.squared	p.value	sigma
0.2808439	0	14.51877

- r.squared = R^2, the proportion of variation in sbp accounted for by the model using dbp.
  - indicates improvement over predicting mean(sbp) for everyone
- p.value = refers to a global F test
  - indicates something about combination of r^2 and sample size
- sigma = residual standard error

glance provides 9 additional summaries for a linear model.

How is the r-squared (r^2)?

R-squared describes the proportion of the variation in `sbp` accounted for by the linear model `m1` using `dbp`.

- R^2 is about 28% (or 0.28) in this case. Is that good?
- Why is this called R-squared? What is the R?

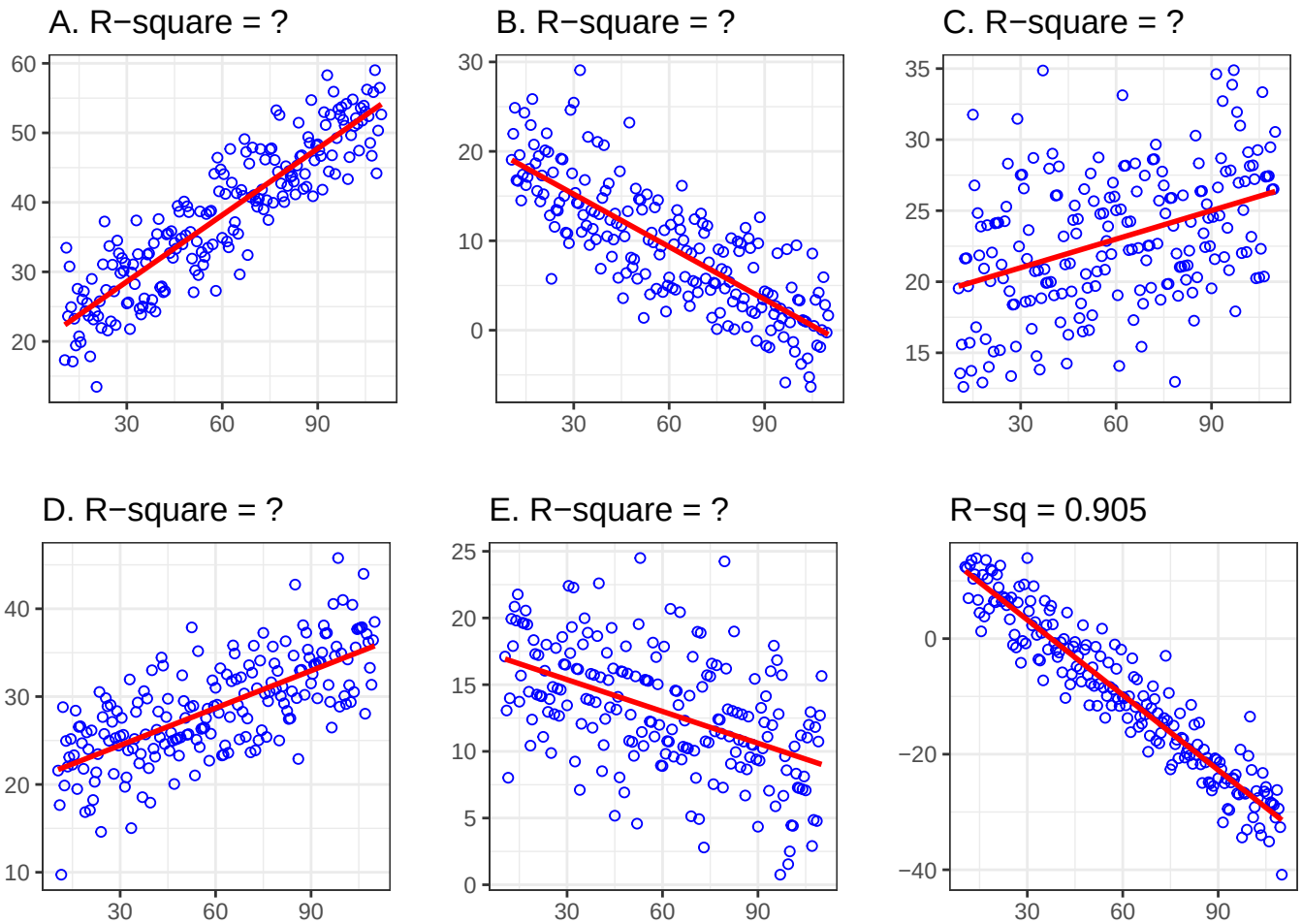
```
nh3_new %$% cor(sbp, dbp)
```

```
[1] 0.5299471
```

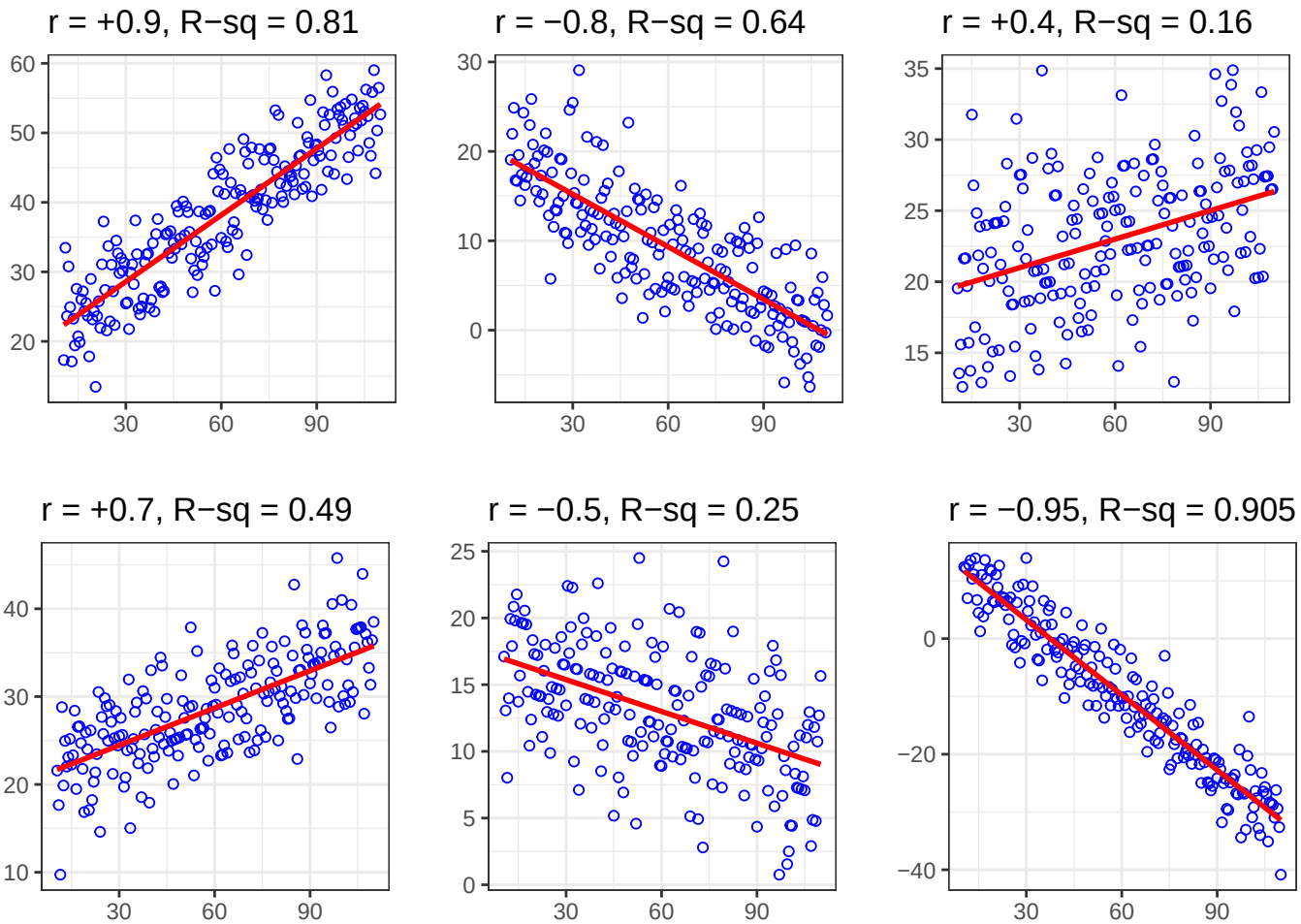
```
nh3_new %$% cor(sbp, dbp)^2
```

```
[1] 0.2808439
```

Can you guess the missing R-squares?



Gaining Insight into what R-square implies



Predict using m1: sbp = 64.27 + 0.795 dbp

Use augment (also from broom) to capture results.

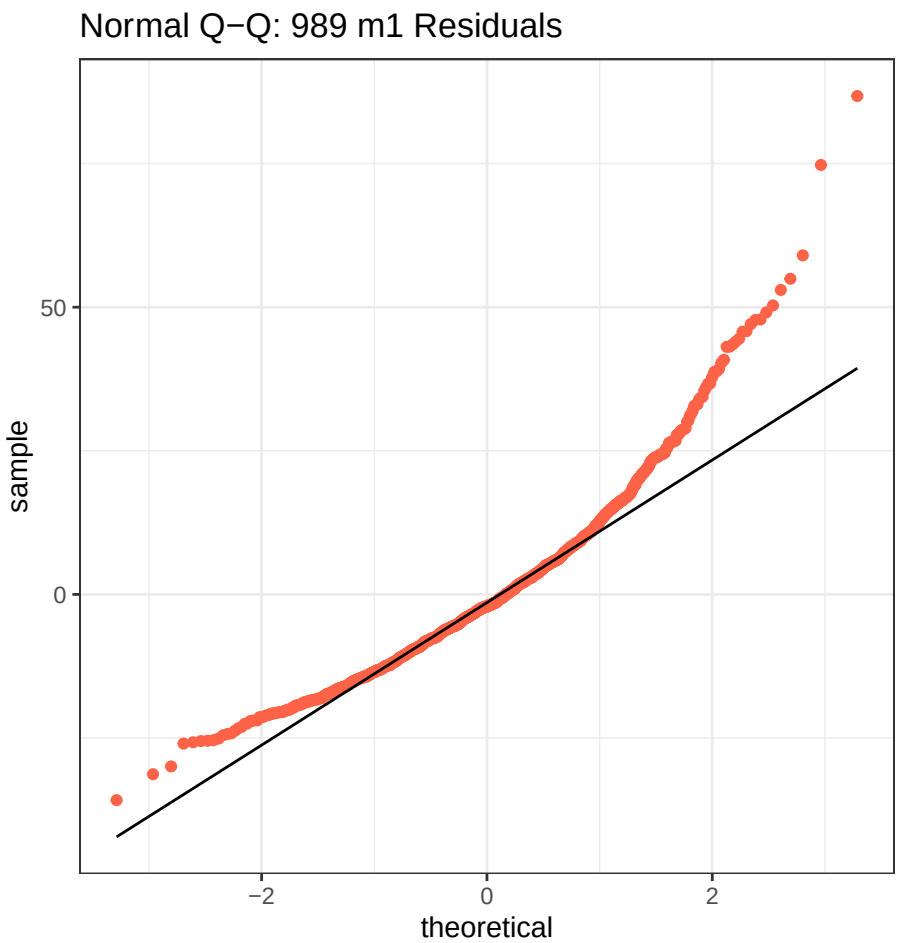
```
m1_insample <- augment(m1, data = nh3_new)

m1_insample %>% select(id, sbp, dbp, .fitted, .resid) %>%
  head(2) %>% kable(digits = 2)
```

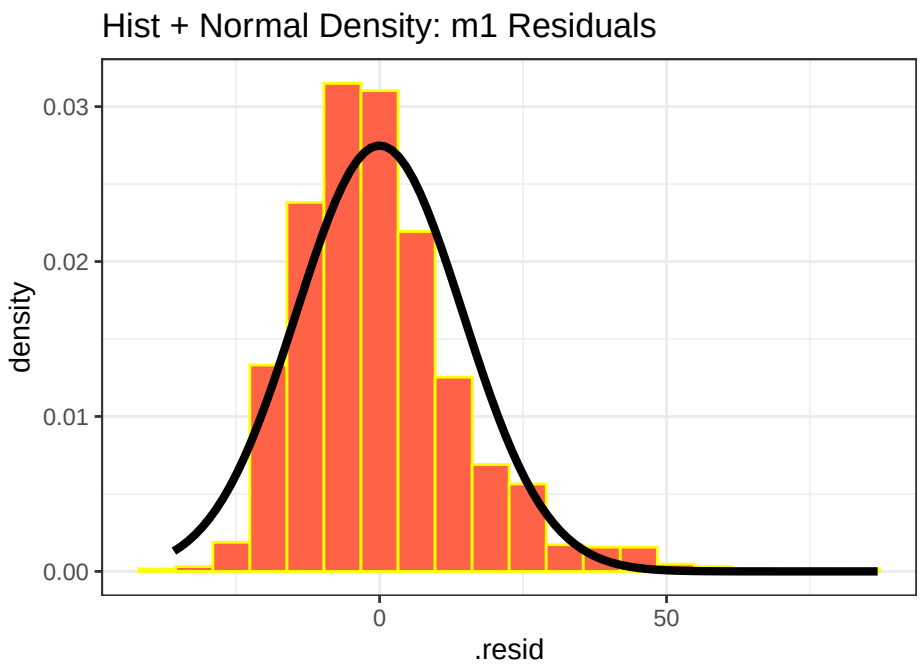
id	sbp	dbp	.fitted	.resid
69036	136	44	99.25	36.75
65956	98	65	115.95	-17.95

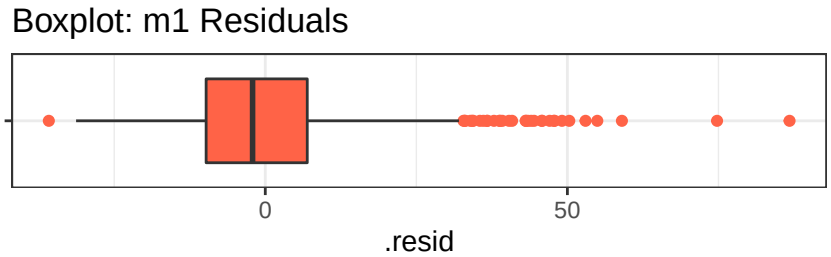
For subject 69036, as an example, we have:

- m1's fitted sbp = 64.27 + 0.795 (44) = 99.25 mm Hg
- **residual** = observed - fitted = 136 - 99.25 = 36.75 mm Hg

Plot residuals from `m1` in our sample (n = 989)

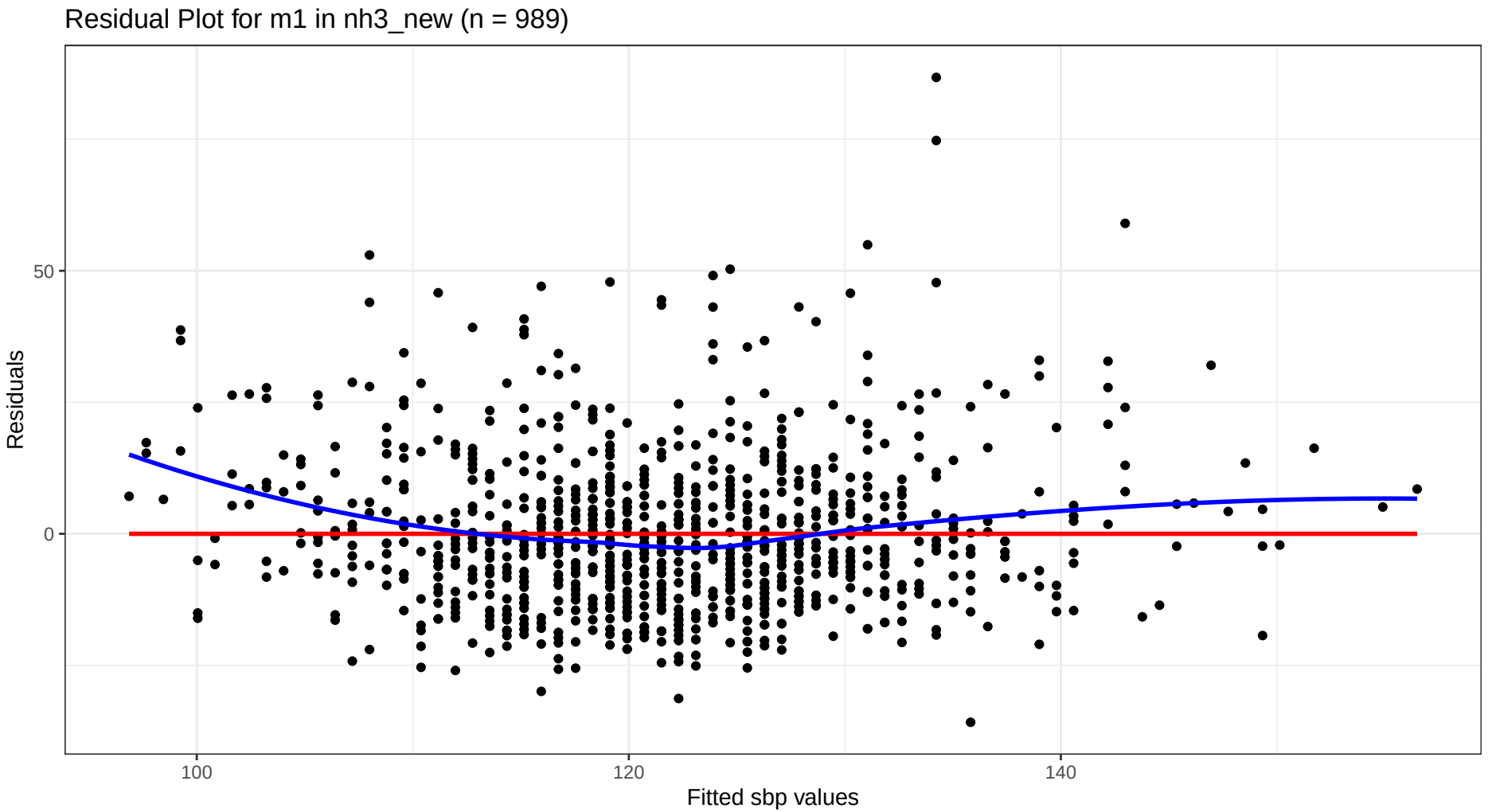






min	Q1	median	Q3	max	mean	sd	n	missing
-35.8	-9.8	-2.1	6.9	86.8	0	14.5	989	0

Plot Residuals vs. Predicted (Fitted) Values



Who else could we make predictions for with m1?

Consider NHANES subjects who we didn't choose for the nh3 sample?

```
nh_deduplicated <- NHANES %>%
    filter(SurveyYr == "2011_12") %>%
    select(ID, SurveyYr, Age, Height, Weight, BMI, Pulse,
           SleepHrsNight, BPSysAve, BPDiaAve, Gender,
           PhysActive, SleepTrouble, Smoke100,
           Race1, HealthGen, Depressed) %>%
    rename(SleepHours = SleepHrsNight, Sex = Gender,
           SBP = BPSysAve, DBP = BPDiaAve) %>%
    filter(Age > 20 & Age < 80) %>%
    drop_na() %>%
    distinct()
```

This nh_deduplicated group is who we sampled from to get nh3.

Identifying those not sampled, but still eligible.

We sampled 1000 observations from a group, and then dropped those with dbp below 40, leaving n = 989. How many people in total would be eligible?

```
nh3_new_eligible <- nh_deduplicated %>%
  clean_names() %>%
  filter(dbp > 39)

dim(nh3_new_eligible)
```

```
[1] 1709   17
```

```
dim(nh3_new)
```

```
[1] 989  17
```

Identify the rest: 1709-989 = 720 not sampled

```
nh3_therest <-
  anti_join(nh3_new_eligible, nh3_new, by = "id")

dim(nh3_therest)
```

```
[1] 720  17
```

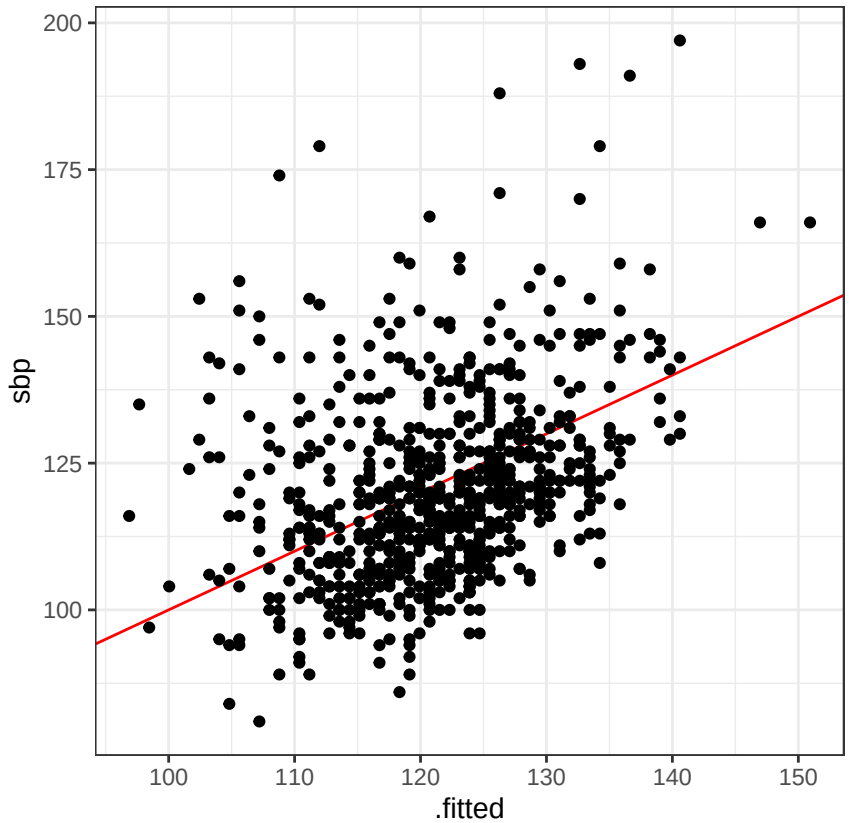
Use model m1 to predict SBP in nh3_therest?

```
new720_nh3 <- augment(m1, newdata = nh3_therest)

new720_nh3 %>% select(id, sbp, dbp, .fitted, .resid) %>%
  head() %>% kable(digits = 2)
```

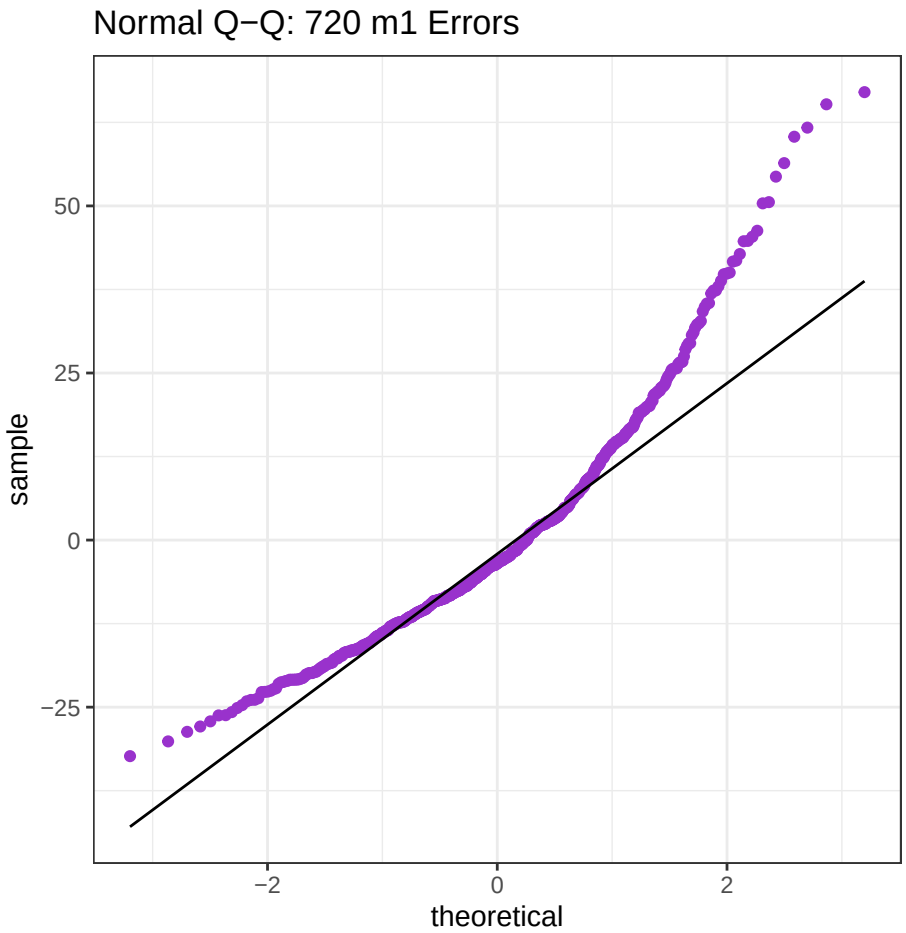
id	sbp	dbp	.fitted	.resid
62172	103	72	121.51	-18.51
62180	107	66	116.74	-9.74
62199	110	65	115.95	-5.95
62205	122	87	133.44	-11.44
62223	105	69	119.13	-14.13
62228	114	74	123.10	-9.10

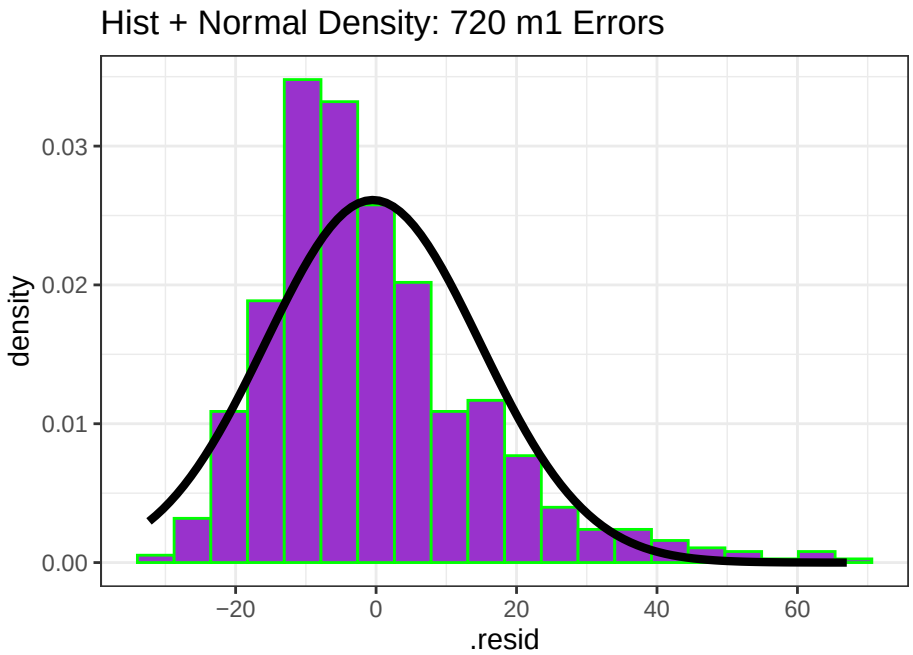
Actual SBP vs. Fitted SBP by `m1` (n = 720)

```
ggplot(new720_nh3, aes(x = .fitted, y = sbp)) +
  geom_abline(slope = 1, intercept = 0, col = "red") +
  geom_point() + theme(aspect.ratio = 1)
```



New Sample (n = 720): `m1` Prediction Errors

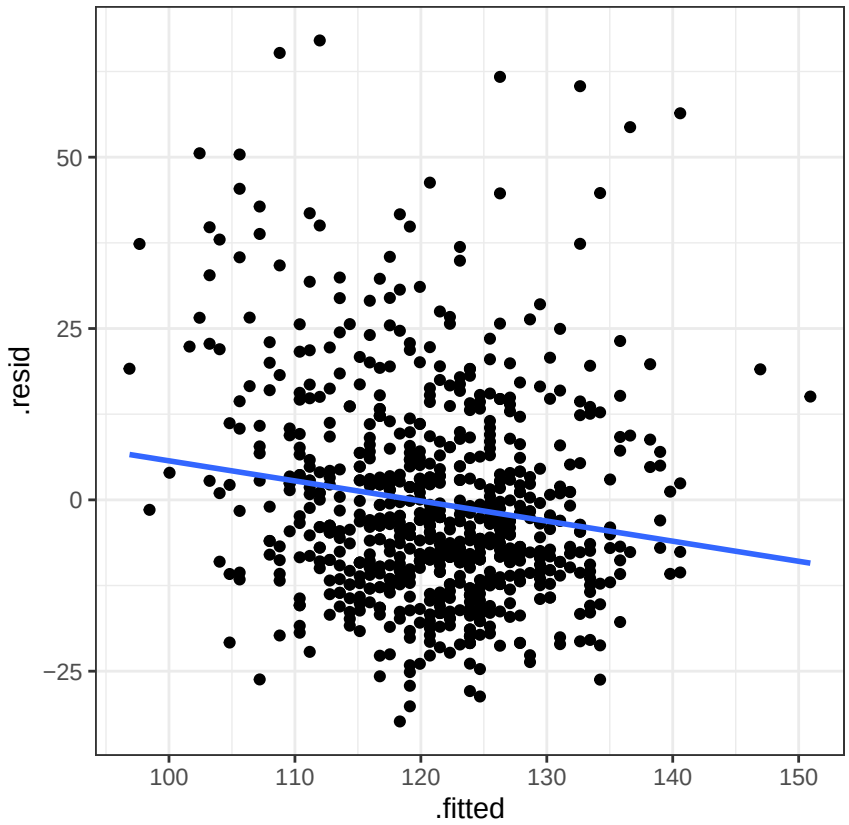




min	Q1	median	Q3	max	mean	sd	n	missing
-32.3	-10.7	-3.5	6.5	67	-0.5	15.3	720	0

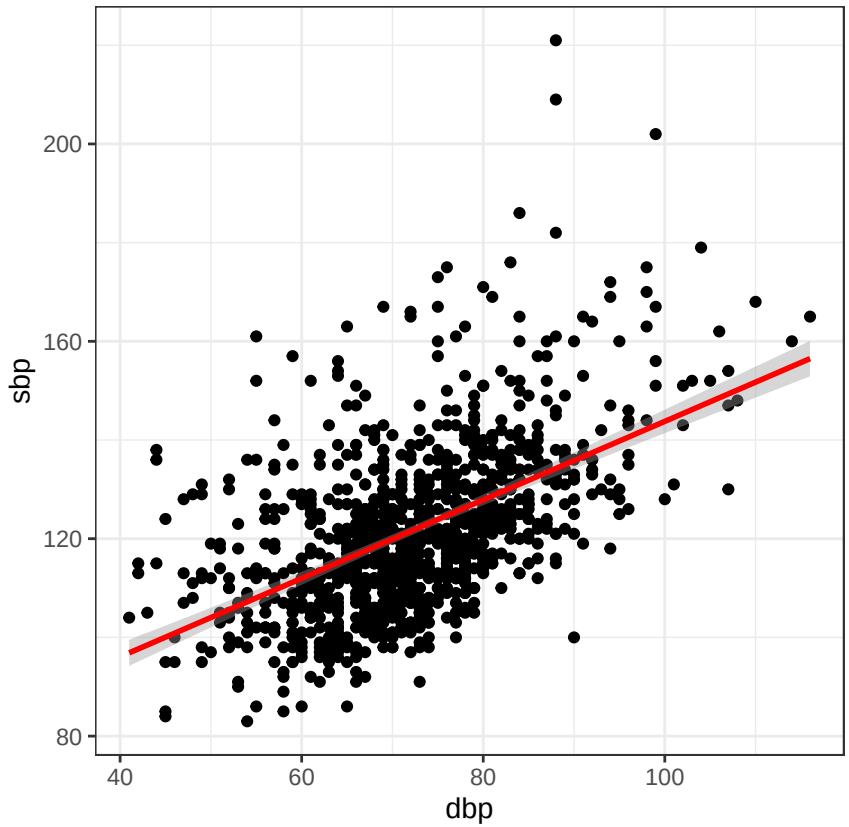
Prediction Errors vs. Fitted SBP (n = 720)

```
ggplot(new720_nh3, aes(x = .fitted, y = .resid)) +
  geom_point() + theme(aspect.ratio = 1) +
  geom_smooth(method = "lm", formula = y ~ x, se = FALSE)
```



Back to sbp and dbp. Does m1 work well here?

```
ggplot(nh3_new, aes(x = dbp, y = sbp)) +
  geom_point() + theme(aspect.ratio = 1) +
  geom_smooth(method = "lm", formula = y ~ x,
              col = "red", se = TRUE)
```



Is this the only linear model R can fit to these data?

Nope.

```
library(rstanarm)
```

```
Loading required package: Rcpp
```

```
This is rstanarm version 2.21.1
```

```
- See https://mc-stan.org/rstanarm/articles/priors for changes
```

```
- Default priors may change, so it's safest to specify priors,
```

```
- For execution on a local, multicore CPU with excess RAM we r
```

```
  options(mc.cores = parallel::detectCores())
```

Fit linear model using stan_glm?

```
m2 <- stan_glm(sbp ~ dbp, data = nh3_new)
```

```
SAMPLING FOR MODEL 'continuous' NOW (CHAIN 1).
Chain 1:
Chain 1: Gradient evaluation took 0 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transiti
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
```

Bayesian fitted linear model for our sbp data

```
print(m2)
```

```
stan_glm
 family:       gaussian [identity]
 formula:      sbp ~ dbp
 observations: 989
 predictors:   2
------
            Median MAD_SD
(Intercept) 64.4    3.0
dbp          0.8    0.0

Auxiliary parameter(s):
      Median MAD_SD
sigma 14.5    0.3

------
```

Is the Bayesian model (with default prior) very different from our `lm` in this situation?

```
coef(m1) # fit with lm
```

```
(Intercept)         dbp
 64.2697456   0.7950429
```

```
coef(m2) # stan_glm with default priors
```

```
(Intercept)         dbp
 64.3647450   0.7937923
```

Note that we could use `tidy` and other `broom` functions for the `lm` model but not (yet) for the `stan_glm` model.

Again, consider sbp and dbp. Does m1 work well?

```
ggplot(nh3_new, aes(x = dbp, y = sbp)) +
  geom_point() + theme(aspect.ratio = 1) +
  geom_smooth(method = "loess", formula = y ~ x,
              col = "blue", se = TRUE)
```