Code ▾

Linear Regression: Sample Model for understanding

libraries

Hide

```
library(ggplot2)
library(tibble)
library(dplyr)
```

```


Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union
```

Hide

```
sim1 <- modelr::sim1
```

Hide

```
sim1
```

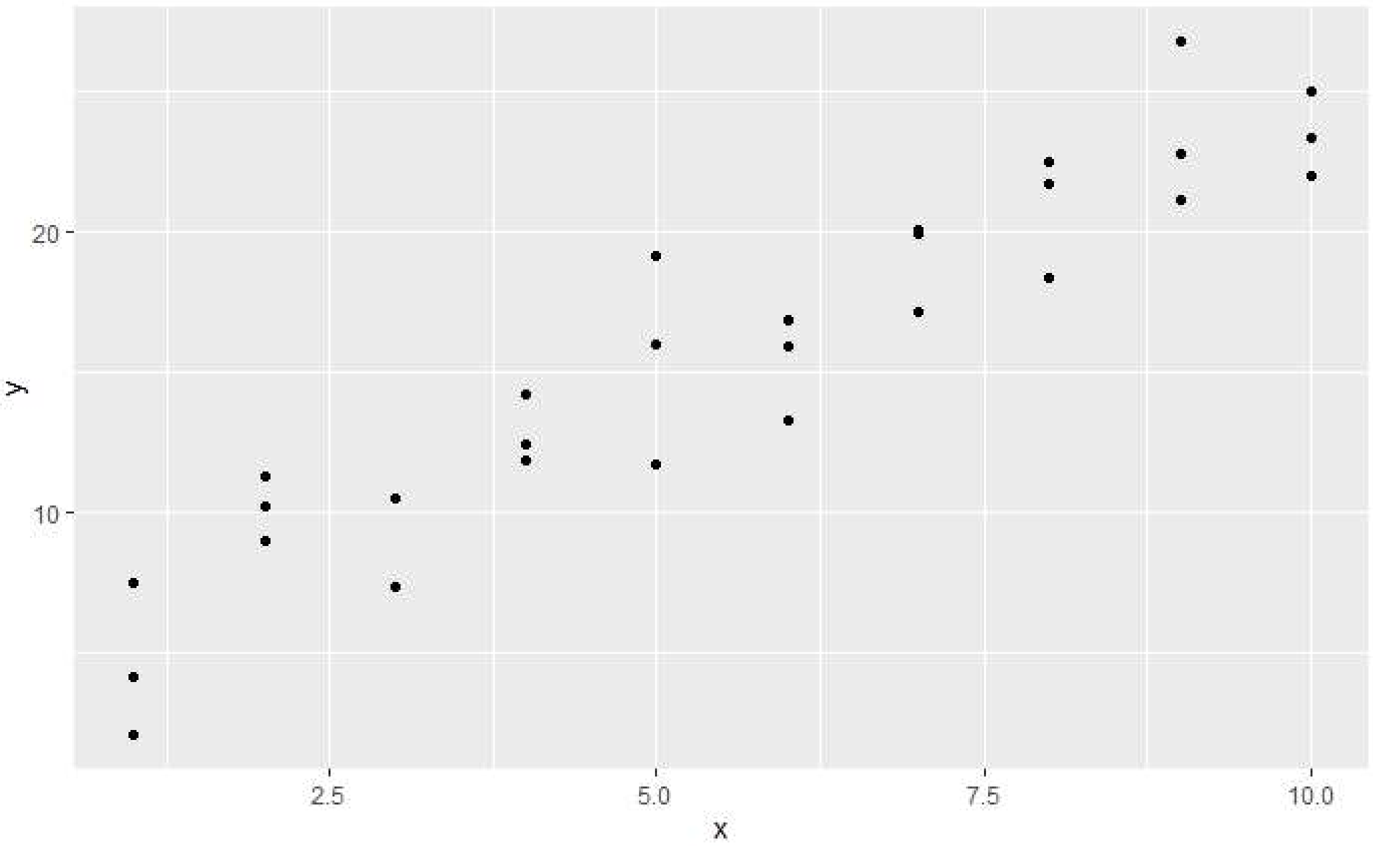
x <int>	y <dbl>
1	4.199913
1	7.510634
1	2.125473
2	8.988857
2	10.243105
2	11.296823
3	7.356365
3	10.505349

x <int>	y <dbl>
3	10.511601
4	12.434589

1-10 of 30 rows Previous **1** 2 3 Next

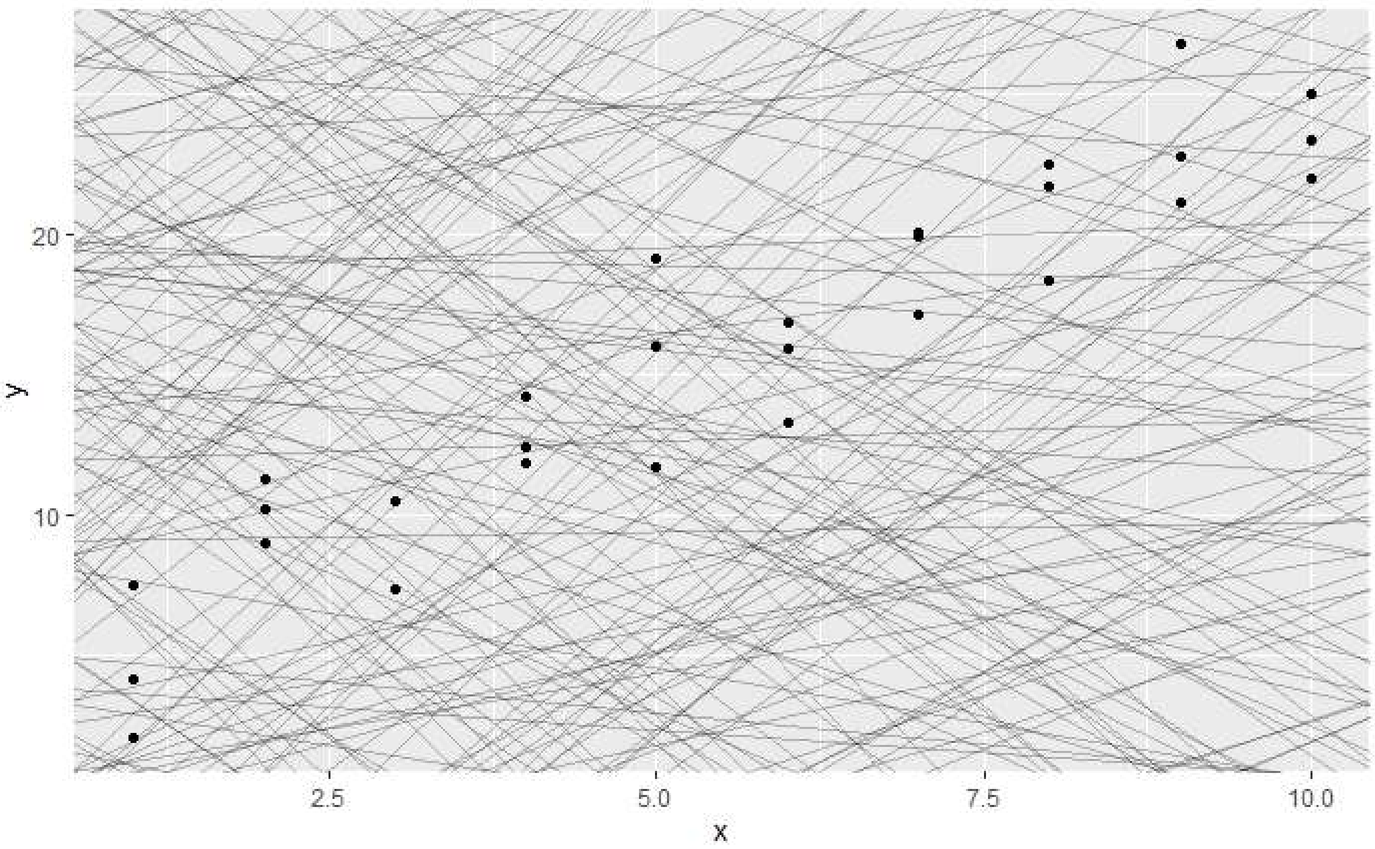
Hide

```
ggplot(sim1, aes(x, y)) +
  geom_point()
```



Hide

```
models <- tibble(
  a1 = runif(250, -20,40),
  a2 = runif(250, -5, 5)
)

ggplot(data = sim1, aes(x, y)) +
  geom_point() +
  geom_abline(aes(intercept=a1, slope=a2), data = models, alpha=1/4)
```



Hide

```
model1 <- function(a, data) {
  a[1] + data$x * a[2]
}

model1(c(7, 1.5), sim1)
```

```
 [1]  8.5  8.5  8.5 10.0 10.0 10.0 11.5 11.5 11.5 13.0 13.0 13.0 14.5 14.5 14.5 16.0 16.0 16.0 1
7.5 17.5 17.5
[22] 19.0 19.0 19.0 20.5 20.5 20.5 22.0 22.0 22.0
```

Root Mean Squared Deviation

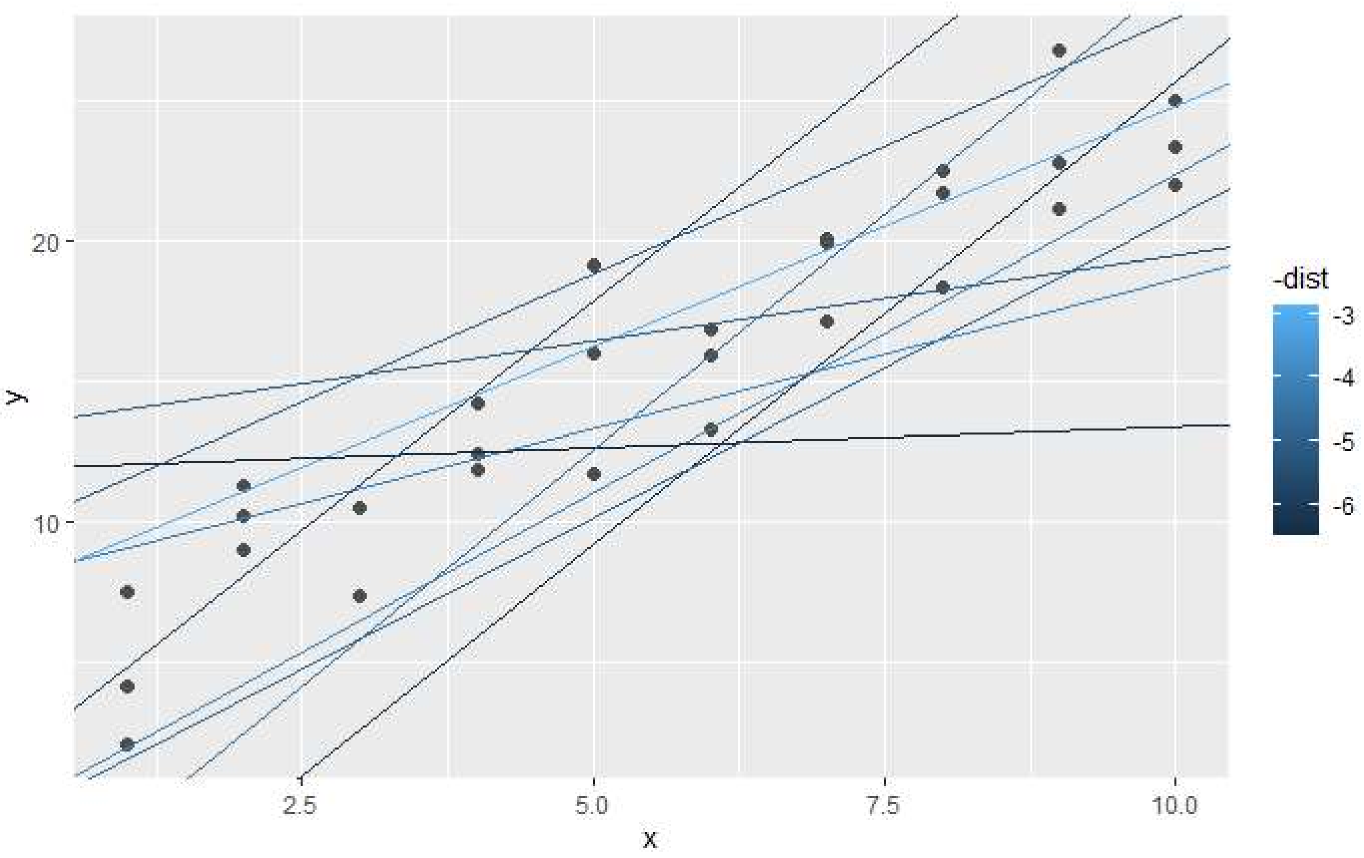
Hide

```
measure_distance <- function(mod, data) {
  diff <- data$y - model1(mod, data)
  sqrt(mean(diff ^ 2))
}

measure_distance(c(7, 1.5), sim1)
```

```
[1] 2.665212
```

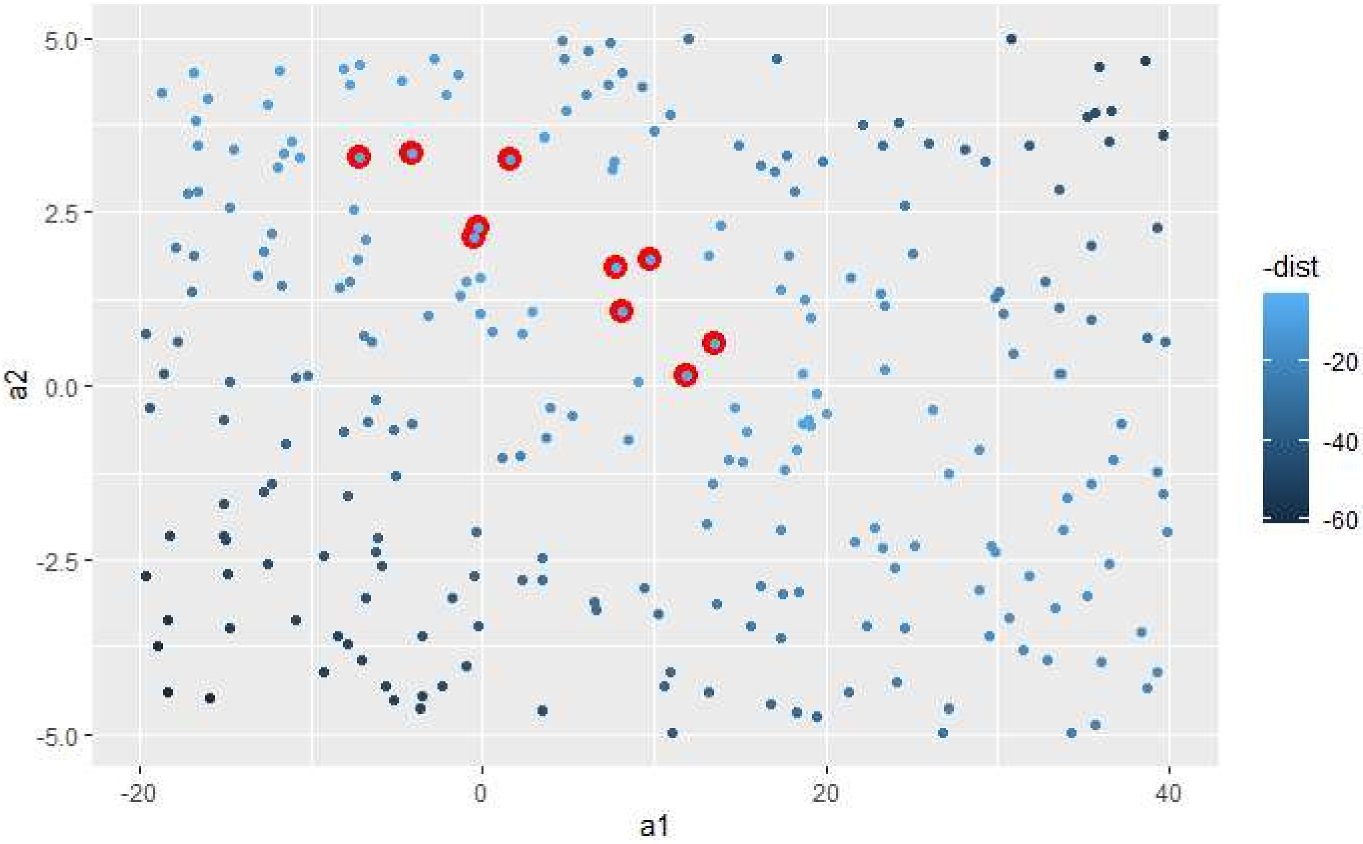
Hide

```
sim1_dist <- function(a1, a2) {
  measure_distance(c(a1,a2), sim1)
}

models <- models %>%
  mutate(dist = purrr::map2_dbl(a1, a2, sim1_dist))

ggplot(sim1, aes(x,y)) +
  geom_point(size=2, color="grey30") +
  geom_abline(
    aes(intercept = a1, slope = a2, color = -dist),
    data = filter(models, rank(dist) <= 10)
  )
```



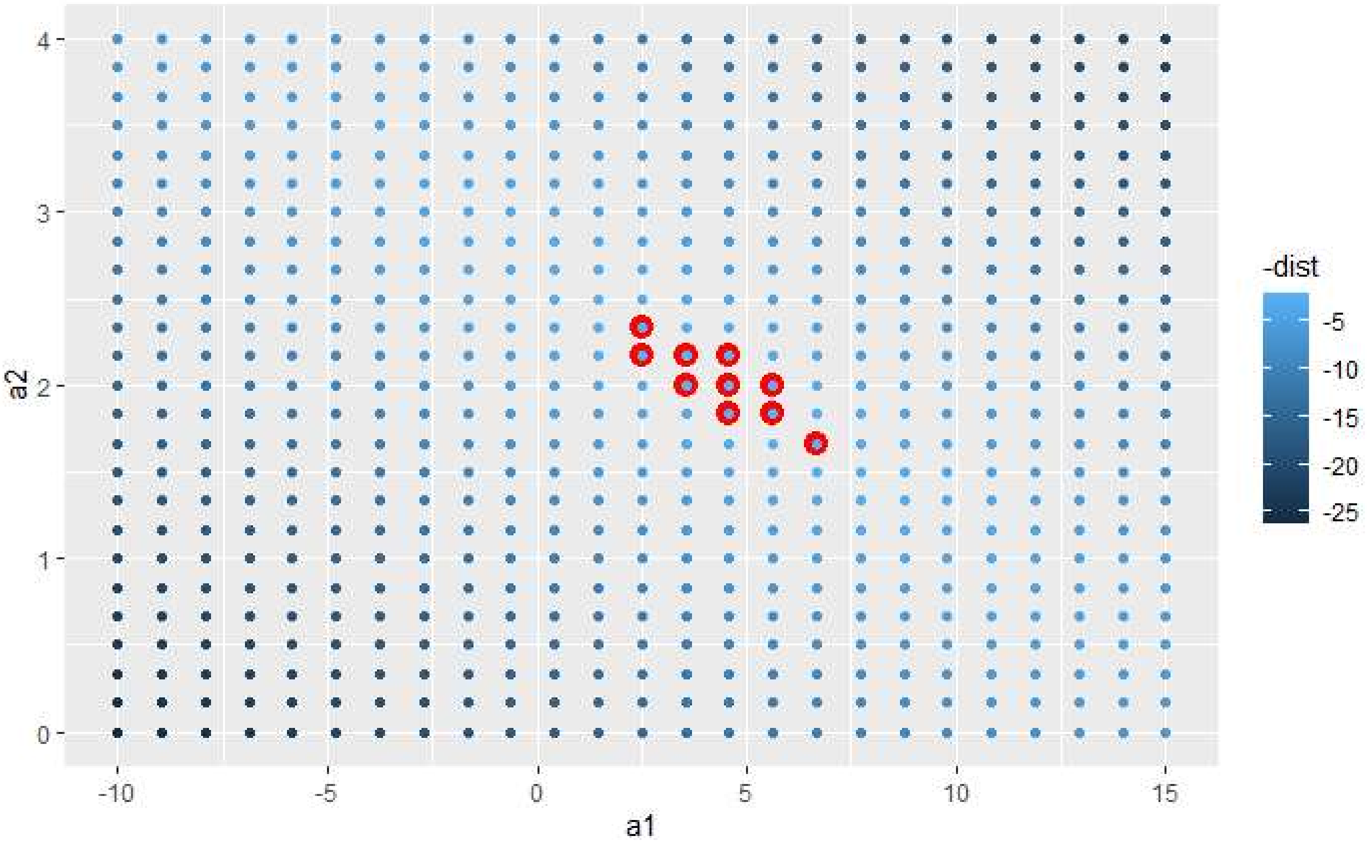
Hide

```
ggplot(models, aes(a1,a2)) +
  geom_point(data = filter(models, rank(dist) <= 10), size=4, color="red")+
  geom_point(aes(color = -dist))
```

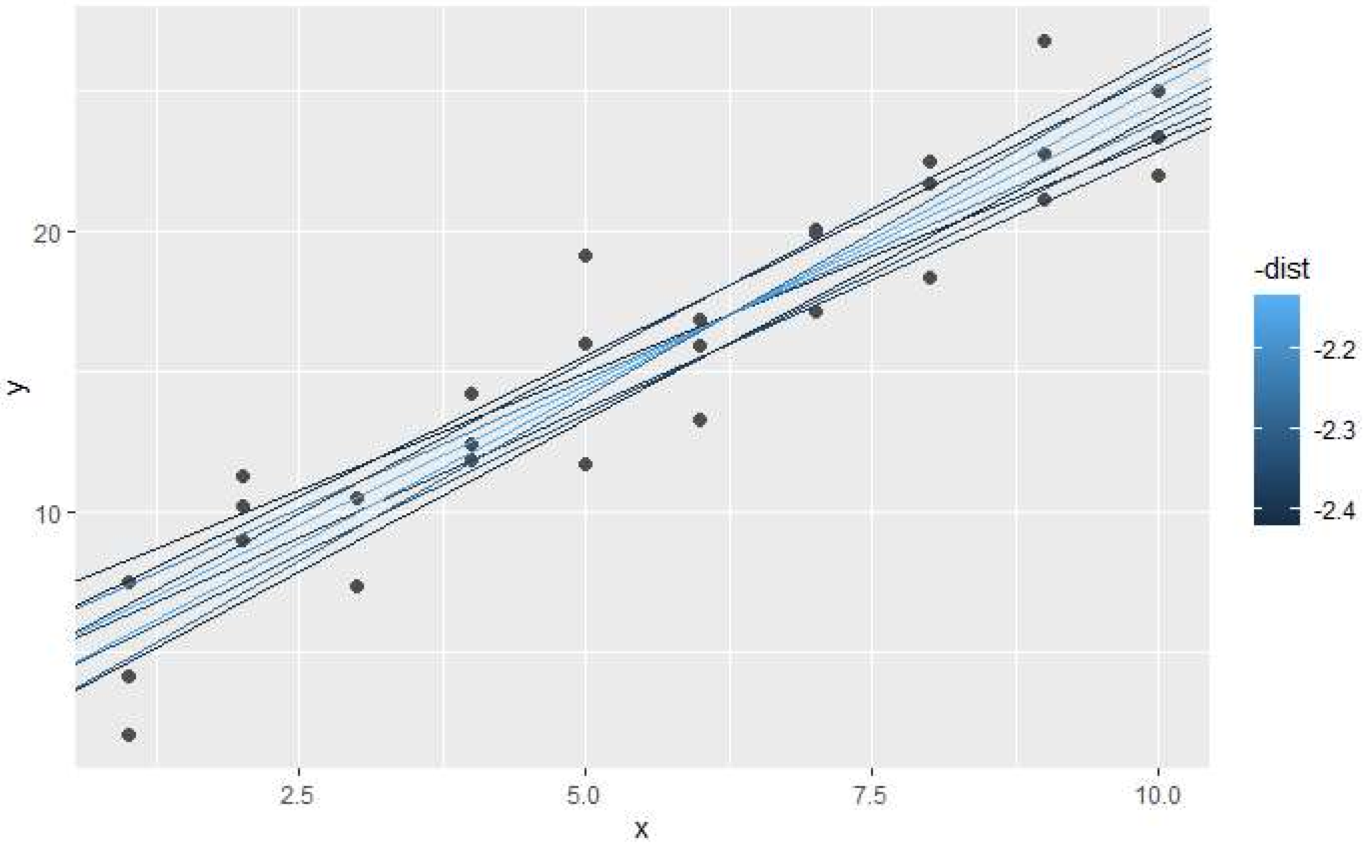


Hide

```
grid <- expand.grid(
  a1 = seq(-10, 15, length =25),
  a2 = seq(0,4, length = 25)
) %>%
  mutate(dist = purrr::map2_dbl(a1,a2, sim1_dist))

grid %>%
  ggplot(aes(a1,a2)) +
  geom_point(data = filter(grid, rank(dist) <= 10), size=4, color="red") +
  geom_point(aes(color = -dist))
```



```
ggplot(sim1, aes(x,y)) +
  geom_point(size =2, color="grey30") +
  geom_abline(
    aes(intercept = a1, slope = a2, color = -dist),
    data = filter(grid, rank(dist) <= 10)
  )
```



Hide

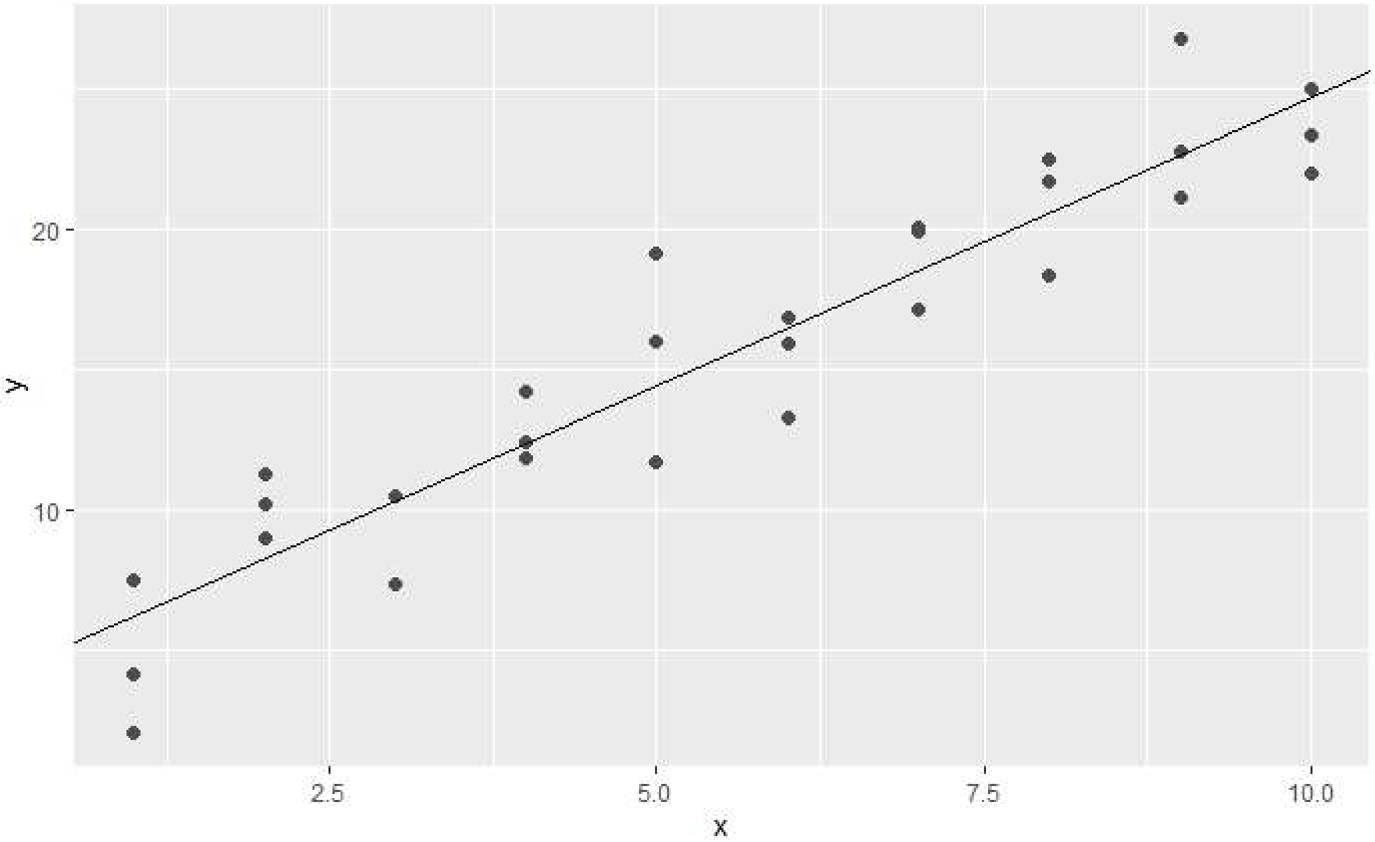
```
best <- optim(c(0,0), measure_distance, data = sim1)
best$par
```

```
[1] 4.222248 2.051204
```

The Best Model from optim package

Hide

```
ggplot(sim1, aes(x,y)) +
  geom_point(size =2, color = "grey30") +
  geom_abline(intercept = best$par[1], slope = best$par[2])
```



Hide

```
sim1_model <- lm(y ~ x, data = sim1)
coef(sim1_model)
```

```
(Intercept)           x 
   4.220822    2.051533 
```

Hide

```
broom::tidy(sim1_model)
```

term <chr>	estimate <dbl>	std.error <dbl>	statistic <dbl>	p.value <dbl>
(Intercept)	4.220822	0.8688261	4.858074	4.088263e-05
x	2.051533	0.1400240	14.651295	1.173451e-14

2 rows

Prediction

Hide

```
library(rsample)
data_split <- initial_split(sim1)
training_data <- training(data_split)
testing_data <- testing(data_split)
```

Hide

```
model <- lm(y ~x, data = training_data)
coef(model)
```

```
(Intercept)           x 
   3.798628    2.148178 
```

Hide

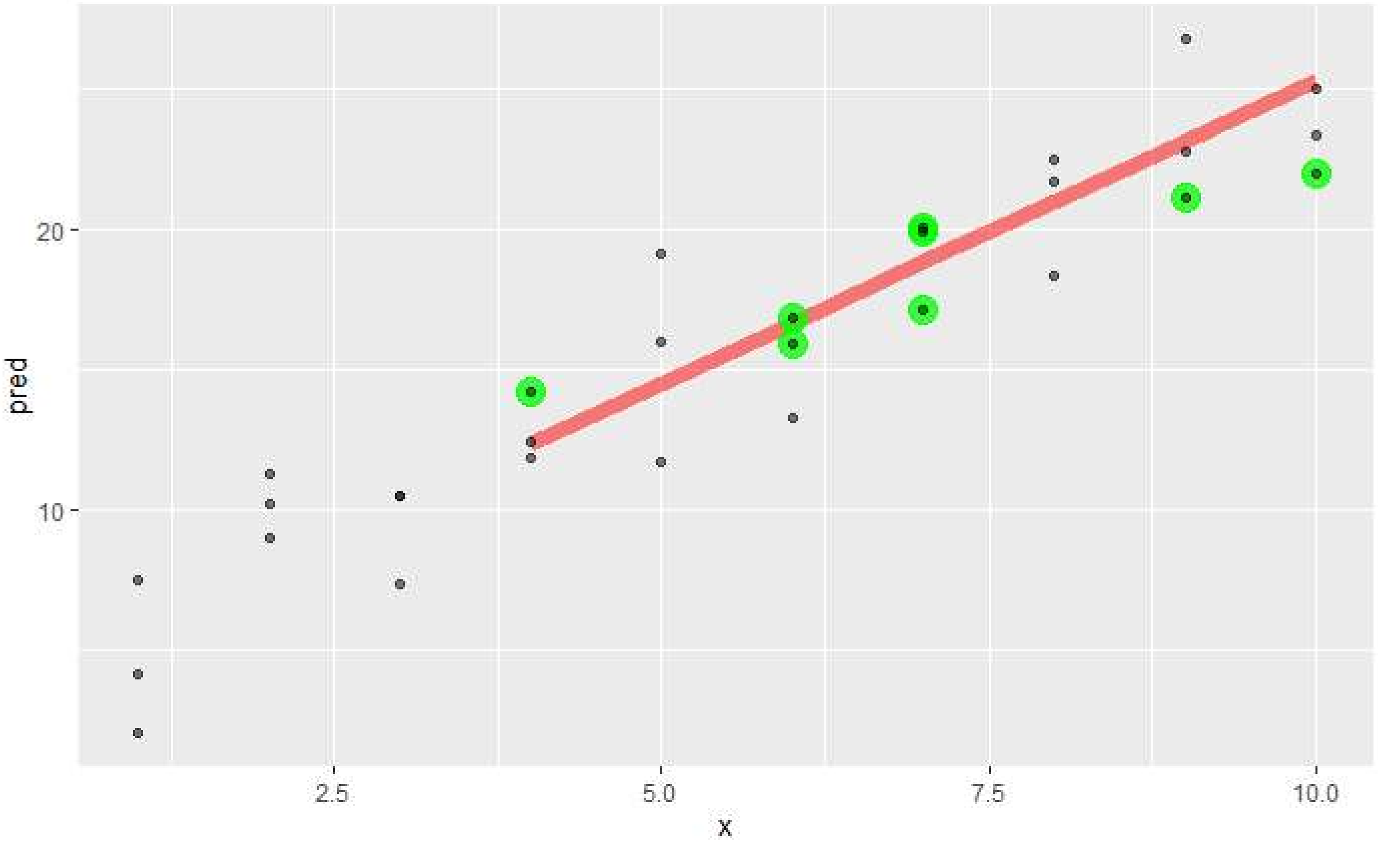
```
prediction <- predict(model, testing_data)
testing_data <- testing_data %>%
  mutate(pred = prediction)
testing_data
```

x	y	pred
<int>	<dbl>	<dbl>
4	14.25796	12.39134
6	15.95597	16.68770
6	16.89480	16.68770
7	20.08599	18.83587
7	17.17185	18.83587
7	19.93631	18.83587
9	21.12831	23.13223
10	21.97520	25.28041

8 rows

Hide

```
ggplot(testing_data) +
  geom_line(aes(x, pred), size = 3, color = "red", alpha = 0.5) +
  geom_point(aes(x,y), size = 5, color ="green", alpha = 3/4) +
  geom_point(data = sim1, aes(x,y), alpha = 0.5)
```



Measure acuracy

Hide

```
yardstick::metrics(testing_data, y, pred)
```

.metric <chr>	**.estimator** <chr>	**.estimate** <dbl>
rmse	standard	1.7516218
rsq	standard	0.8440479
mae	standard	1.5161443

3 rows