Lab6

Exercises

Irimie Fabio

Contents

Exercise 1

A

Compare the PDF's and the CDF's of the following Uniform continuous random variables (hint: see page 3 of the slides).

Calculate: 1. $X1 \sim U(0, 1)$

```
a <- 0
b <- 1
seq <- seq(a - 0.5, b + 0.5, 0.01)
x1 <- data.frame(
  x = seq,
  pdf = dunif(seq, a, b),
  cdf = punif(seq, a, b)
)
```

2. $X2 \sim U(-3, 2)$

```r
a <- -3
b <- 2
seq <- seq(a - 0.5, b + 0.5, 0.01)
x2 <- data.frame(
  x = seq,
  pdf = dunif(seq, a, b),
  cdf = punif(seq, a, b)
)
```

3. $X3 \sim U(2, 4)$

```r
a <- 2
b <- 4
seq <- seq(a - 0.5, b + 0.5, 0.01)
x3 <- data.frame(
  x = seq,
  pdf = dunif(seq, a, b),
  cdf = punif(seq, a, b)
)
```

4. $X4 \sim U(0.8, 2.5)$

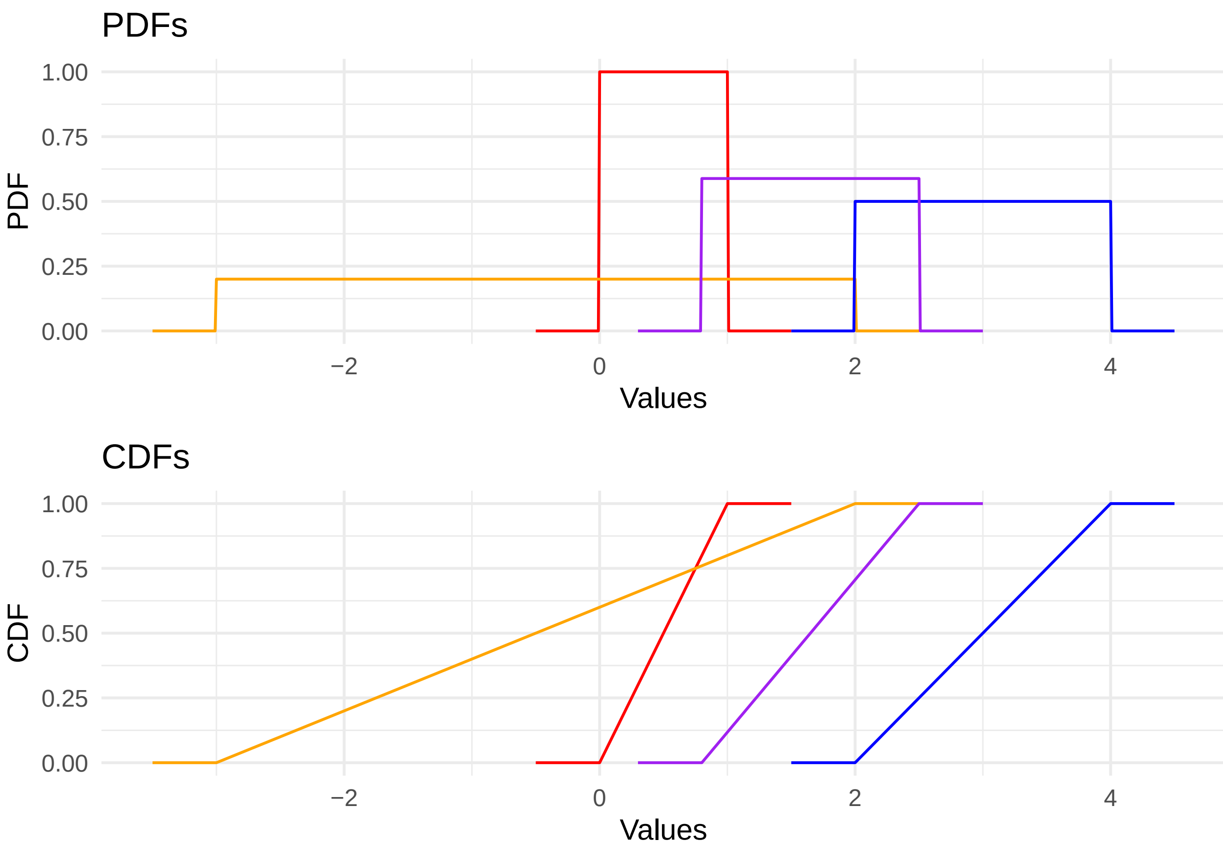
```r
a <- 0.8
b <- 2.5
seq <- seq(a - 0.5, b + 0.5, 0.01)
x4 <- data.frame(
  x = seq,
  pdf = dunif(seq, a, b),
  cdf = punif(seq, a, b)
)
```

B

Create a new figure with two vertical panels: first row PDFs, second row CDFs (hint: use the library cowplot and the plot_grid() function to arrange multiple ggplot objects).

```r
library(ggplot2)
library(cowplot)

pdfs <- ggplot() +
  geom_line(data = x1, aes(x = x, y = pdf), color = "red") +
  geom_line(data = x2, aes(x = x, y = pdf), color = "orange") +
  geom_line(data = x3, aes(x = x, y = pdf), color = "blue") +
  geom_line(data = x4, aes(x = x, y = pdf), color = "purple") +
  labs(title = "PDFs", x = "Values", y = "PDF") +
  theme_minimal()


cdfs <- ggplot() +
  geom_line(data = x1, aes(x = x, y = cdf), color = "red") +
  geom_line(data = x2, aes(x = x, y = cdf), color = "orange") +
  geom_line(data = x3, aes(x = x, y = cdf), color = "blue") +
  geom_line(data = x4, aes(x = x, y = cdf), color = "purple") +
  labs(title = "CDFs", x = "Values", y = "CDF") +
  theme_minimal()
```

```r
plot_grid(pdfs, cdfs, ncol = 1)
```



C

Determine the mean and variance of each random variable.

```r
x1_mean <- mean(runif(1000000, 0, 1))
x1_var <- var(runif(1000000, 0, 1))

x2_mean <- mean(runif(1000000, -3, 2))
x2_var <- var(runif(1000000, -3, 2))

x3_mean <- mean(runif(1000000, 2, 4))
x3_var <- var(runif(1000000, 2, 4))

x4_mean <- mean(runif(1000000, 0.8, 2.5))
x4_var <- var(runif(1000000, 0.8, 2.5))

table <- data.frame(
  x = c("X1", "X2", "X3", "X4"),
  mean = c(x1_mean, x2_mean, x3_mean, x4_mean),
  var = c(x1_var, x2_var, x3_var, x4_var)
)
table
```

```
##    x       mean        var
## 1 X1  0.5001694 0.08342179
## 2 X2 -0.4985534 2.08264539
```

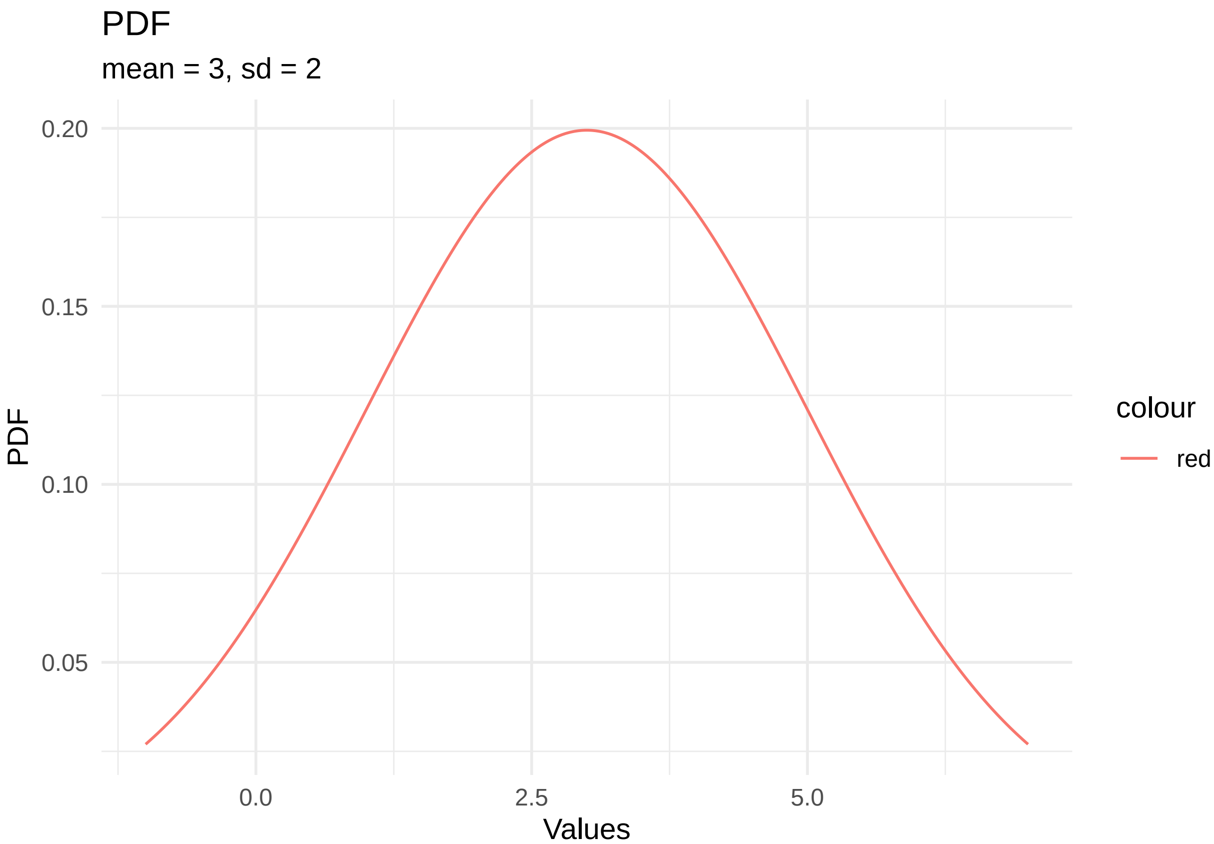
```
## 3 X3  3.0007472 0.33340818
## 4 X4  1.6501204 0.24069993
```

Exercise 2

A

Determine and plot the values of the PDF for a normal distribution with mean 3 and sd 2 for x values between -1 and 7.

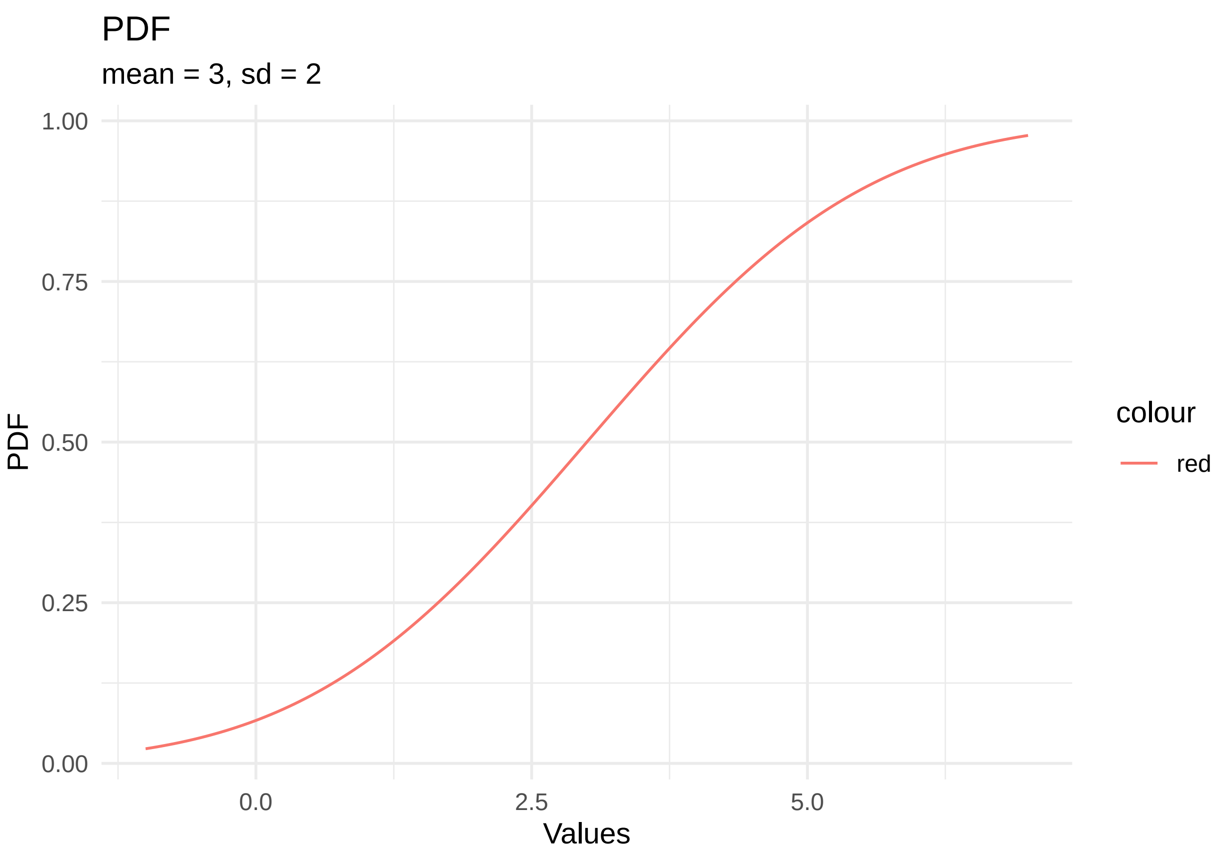
```r
mu <- 3
sd <- 2
seq <- seq(-1, 7, 0.01)
normal <- data.frame(
  x = seq,
  pdf = dnorm(seq, mu, sd)
)

ggplot(normal, aes(x = x, y = pdf)) +
  geom_line(aes(y = pdf, color = "red")) +
  labs(
    title = "PDF",
    subtitle = "mean = 3, sd = 2",
    x = "Values",
    y = "PDF"
  ) +
  theme_minimal()
```



B

Determine and plot the corresponding values of the CDF

```r
mu <- 3
sd <- 2
seq <- seq(-1, 7, 0.01)
normal <- data.frame(
  x = seq,
  pdf = pnorm(seq, mu, sd)
)

ggplot(normal, aes(x = x, y = pdf)) +
  geom_line(aes(y = pdf, color = "red")) +
  labs(
    title = "PDF",
    subtitle = "mean = 3, sd = 2",
    x = "Values",
    y = "PDF"
  ) +
  theme_minimal()
```



C

Determine for what values of x the value of the CDF equals 0.025 and 0.975 (hint: remember all the prefixes: p, d, r, q. See Lab 4, slide 2).

```r
qnorm(0.025, mu, sd)
```

```
## [1] -0.919928
```

```r
qnorm(0.975, mu, sd)
```

```
## [1] 6.919928
```

Exercise 3

A

Create a set of values ranging from xmin to xmax.

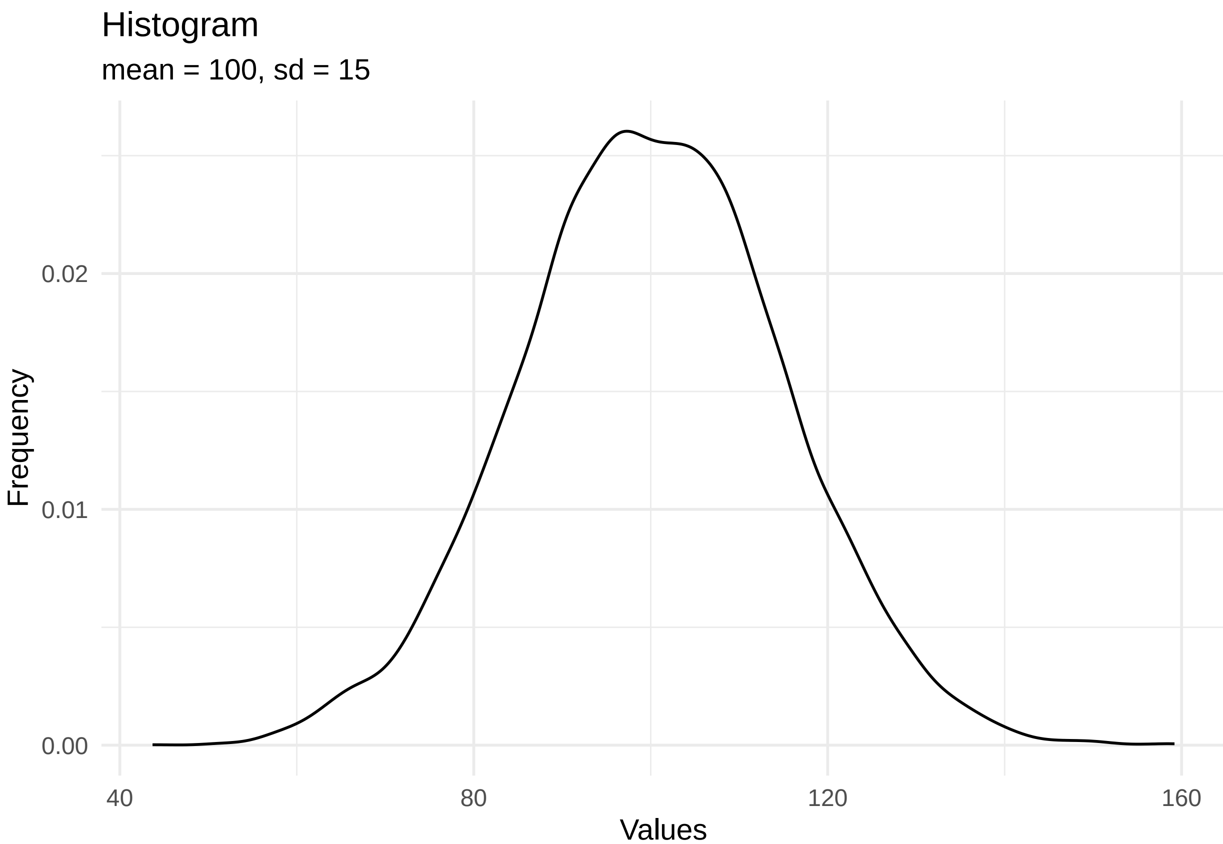
```r
mu <- 100 # the mean
sigma <- 15 # the standard deviation
xmin <- 70 # minumum x value for pdf and cdf plots
xmax <- 130 # maximum x value for pdf and cdf plots
n <- 100 # number of points of pdf and cdf plots
k <- 10000 # number of random draws (for histogram)

seq <- seq(xmin, xmax, length = n)
seq
```

```
##   [1]  70.00000  70.60606  71.21212  71.81818  72.42424  73.03030
##   [7]  73.63636  74.24242  74.84848  75.45455  76.06061  76.66667
##  [13]  77.27273  77.87879  78.48485  79.09091  79.69697  80.30303
##  [19]  80.90909  81.51515  82.12121  82.72727  83.33333  83.93939
##  [25]  84.54545  85.15152  85.75758  86.36364  86.96970  87.57576
##  [31]  88.18182  88.78788  89.39394  90.00000  90.60606  91.21212
##  [37]  91.81818  92.42424  93.03030  93.63636  94.24242  94.84848
##  [43]  95.45455  96.06061  96.66667  97.27273  97.87879  98.48485
##  [49]  99.09091  99.69697 100.30303 100.90909 101.51515 102.12121
##  [55] 102.72727 103.33333 103.93939 104.54545 105.15152 105.75758
##  [61] 106.36364 106.96970 107.57576 108.18182 108.78788 109.39394
##  [67] 110.00000 110.60606 111.21212 111.81818 112.42424 113.03030
##  [73] 113.63636 114.24242 114.84848 115.45455 116.06061 116.66667
##  [79] 117.27273 117.87879 118.48485 119.09091 119.69697 120.30303
##  [85] 120.90909 121.51515 122.12121 122.72727 123.33333 123.93939
##  [91] 124.54545 125.15152 125.75758 126.36364 126.96970 127.57576
##  [97] 128.18182 128.78788 129.39394 130.00000
```

B

Draw k random numbers from a $N(\mu, \sigma^2)$ distribution.

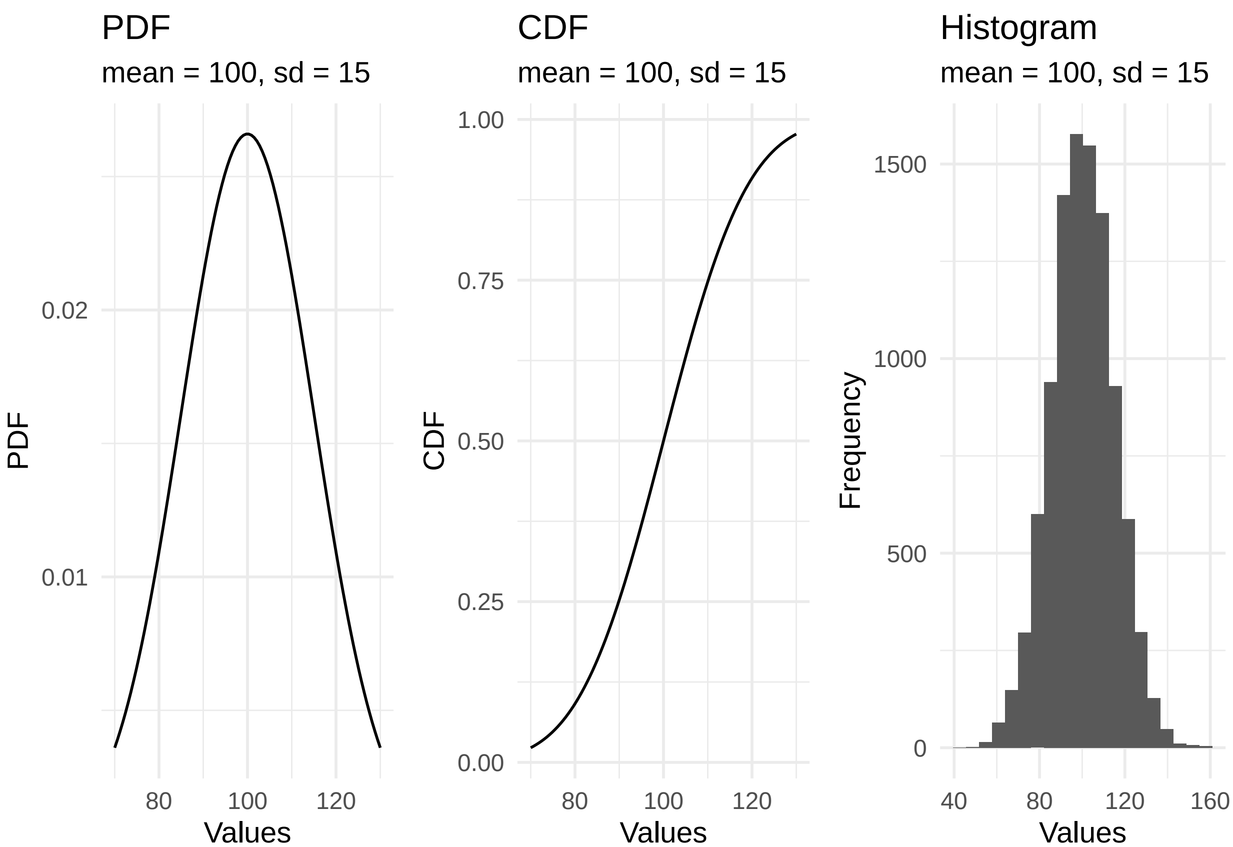
```r
x <- rnorm(k, mu, sigma)

ggplot(data.frame(x = x), aes(x = x)) +
  geom_line(aes(y = ..density..), stat = "density") +
  labs(
    title = "Histogram",
    subtitle = "mean = 100, sd = 15",
    x = "Values",
    y = "Frequency"
  ) +
  theme_minimal()
```



C Create a new figure with 3 panels: PDF, CDF, and histogram of random values with 20 bins.

```
pdf <- data.frame(
  x = seq,
  pdf = dnorm(seq, mu, sigma)
)

cdf <- data.frame(
  x = seq,
  cdf = pnorm(seq, mu, sigma)
)

histogram <- data.frame(
  x = x
)

pdf_plot <- ggplot(pdf, aes(x = x, y = pdf)) +
  geom_line(aes(y = pdf)) +
  labs(
    title = "PDF",
    subtitle = "mean = 100, sd = 15",
    x = "Values",
    y = "PDF"
  ) +
  theme_minimal()

cdf_plot <- ggplot(cdf, aes(x = x, y = cdf)) +
  geom_line(aes(y = cdf)) +
```

```r
  labs(
    title = "CDF",
    subtitle = "mean = 100, sd = 15",
    x = "Values",
    y = "CDF"
  ) +
  theme_minimal()

histogram_plot <- ggplot(histogram, aes(x = x)) +
  geom_histogram(bins = 20) +
  labs(
    title = "Histogram",
    subtitle = "mean = 100, sd = 15",
    x = "Values",
    y = "Frequency"
  ) +
  theme_minimal()

plot_grid(pdf_plot, cdf_plot, histogram_plot, nrow = 1)
```



Exercise 4

The number of visits per day to a website follows a Poisson distribution with a mean of 120 visits per day. The duration of each visit follows an exponential distribution with a mean time of 5 minutes. The percentage of visitors who purchase something is 20%, and the bills follow a normal distribution with a mean of 50€ and a standard deviation of 10€.

A

Calculate the probability that at least 120 visits are made in a day.

```r
lambda <- 120
ppois(119, lambda, lower.tail = FALSE)
```

```
## [1] 0.51214
```

B

Calculate the expected value and variance of the average duration of a visit.

```r
lambda <- 1 / 5
mean <- 1 / lambda
var <- 1 / lambda^2
mean
```

```
## [1] 5
```

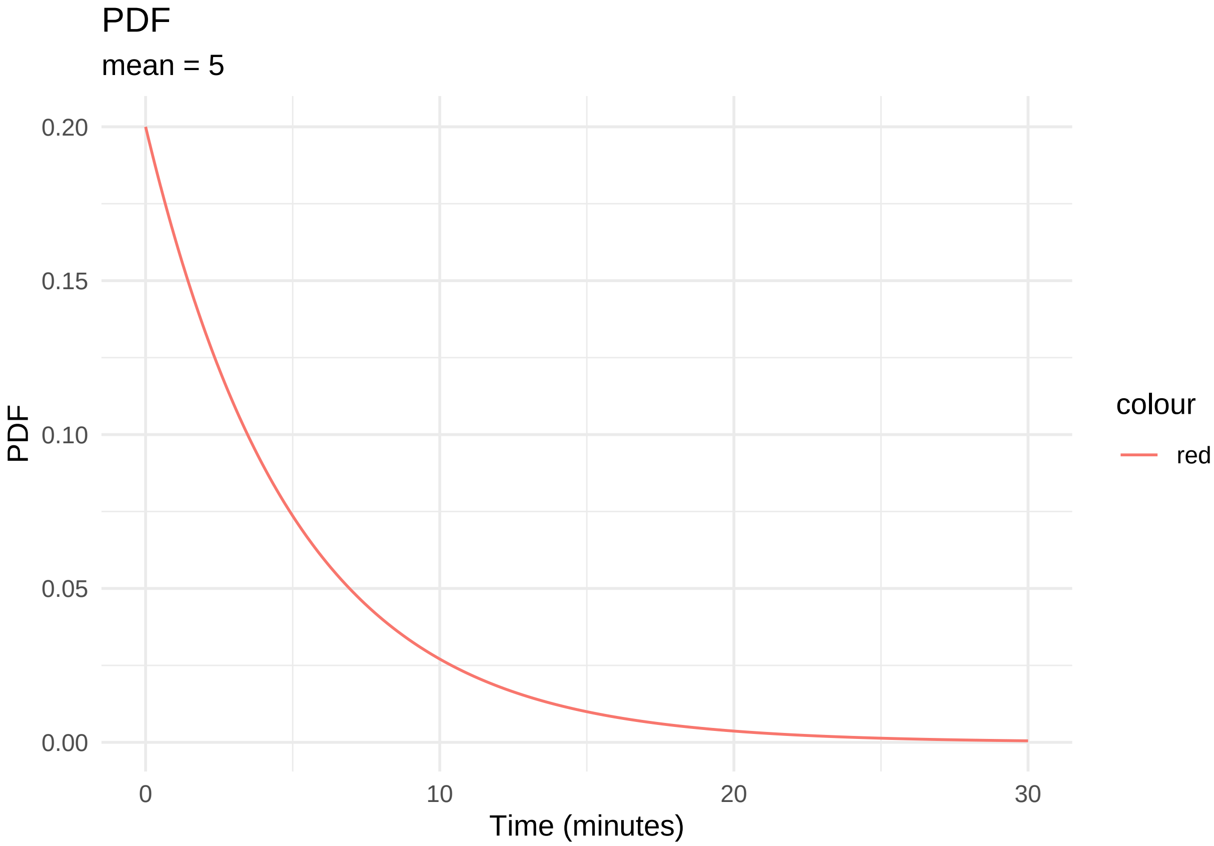
```r
var
```

```
## [1] 25
```

C

Draw the PDF for the random variable “duration of a visit”.

```r
seq <- seq(0, 30, 0.01)
exponential <- data.frame(
  x = seq,
  pdf = dexp(seq, lambda)
)

ggplot(exponential, aes(x = x, y = pdf)) +
  geom_line(aes(y = pdf, color = "red")) +
  labs(
    title = "PDF",
    subtitle = "mean = 5",
    x = "Time (minutes)",
    y = "PDF"
  ) +
  theme_minimal()
```



D

Calculate the probability that a visit lasts from 5 to 10 minutes.

```
pexp(10, lambda) - pexp(5, lambda)
```

```
## [1] 0.2325442
```

E

Given that exactly 100 visits were made in a day, what is the probability to observe at least 20 purchases?

```
lambda <- 0.2 * 100
ppois(19, lambda, lower.tail = FALSE)
```

```
## [1] 0.5297427
```

F

Compute the probabilities to observe a purchase between 40 and 60, 30 and 70, 20 and 80.

```
mu <- 50
sigma <- 10
pnorm(60, mu, sigma) - pnorm(40, mu, sigma)
```

```
## [1] 0.6826895
```

```
pnorm(70, mu, sigma) - pnorm(30, mu, sigma)
```

```
## [1] 0.9544997
```

```r
pnorm(80, mu, sigma) - pnorm(20, mu, sigma)
```

```
## [1] 0.9973002
```

G

Simulate 1000 average days, compute the average daily revenue and compare it to the expected revenue.

```r
n <- 1000
lambda_visits <- 120
lambda_duration <- 1 / 5
purchases <- 0.2
mu_bills <- 50
sigma_bills <- 10

visits <- rpois(n, lambda_visits)
durations <- rexp(n, lambda_duration)
purchases <- rbinom(n, visits, purchases)
bills <- rnorm(n, mu_bills, sigma_bills)
revenue <- purchases * bills

daily_revenue <- mean(revenue)
expected_revenue <- lambda_visits * purchases * mu_bills

daily_revenue
```

```
## [1] 1191.118
```

```r
mean(expected_revenue)
```

```
## [1] 144300
```