

The [Goldilocks principle](#) has its origins in a children's story about a girl who tastes the bowls of porridge left by three bears. She prefers the one that is neither too hot nor too cold, but is just right.

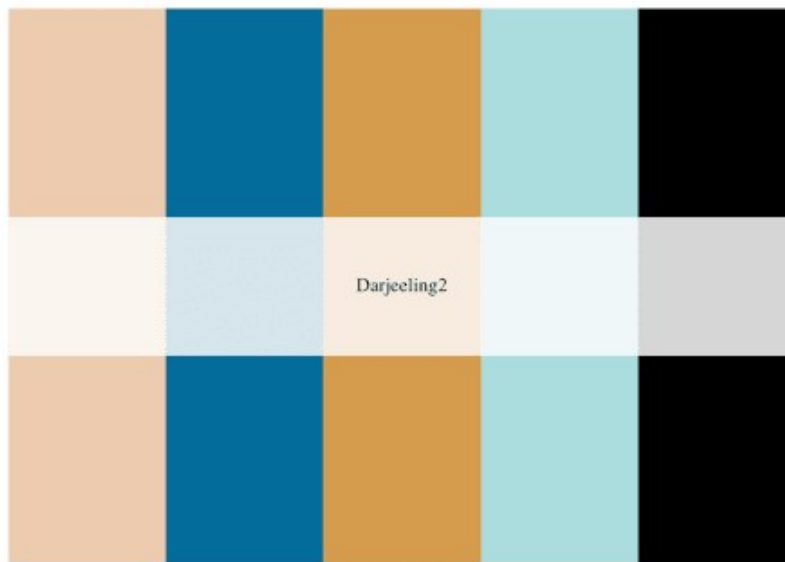
When it comes to investing in stocks, how many is “just right”?

```
library(tidyverse)
library(rebus)
library(wesanderson)
library(kableExtra)
library(scales)
library(truncnorm)
```

I'll use this palette.

```
theme_set(theme_bw())

(cols <- wes_palette(name = "Darjeeling2"))
```



Suppose the [average stock market return](#) is around 10%. And you do extensive research, burning the midnight oil, poring over stock fundamentals. Or perhaps you develop a cool machine learning model. And you arrive at a list of 50 promising stocks you feel confident would, on average, deliver well-above-market returns.

I'll create some randomly made up stocks with an average return close to 40%. Some will tank due to events one could not foresee; I'll allow some to lose up to 20%. Similarly, some could generate exceptional returns.

```
set.seed(123)

stock_data <- tibble(
  stock = chartr("0123456789", "abcdefghij", sample(50)),
  return = rtruncnorm(50, a = -0.2, mean = 0.4, sd = 0.5)
)

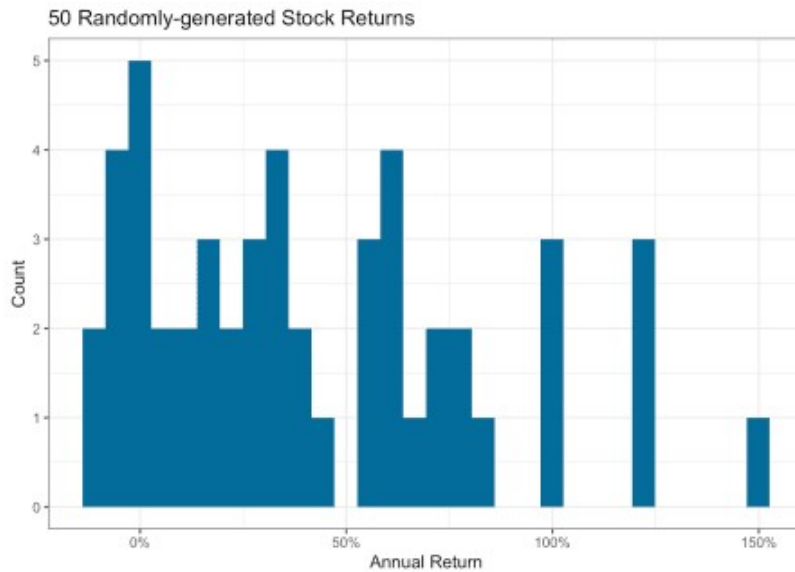
mean(stock_data$return)

## [1] 0.4200244
```

Here's the resultant distribution I'll use to assess the impact of portfolio size. Stock markets are fairly close to a normal distribution, albeit with fatter tails due to a few extreme outcomes.

```
stock_data %>%
```

```
ggplot(aes(return)) +
  geom_histogram(fill = cols[2]) +
  scale_x_continuous(labels = percent_format()) +
  labs(title = "50 Randomly-generated Stock Returns",
       x = "Annual Return", y = "Count")
```



Now suppose you share 2 stocks, selected at random, with 1,000 of your social network friends (selecting a different pair of stocks for each friend). Will they all still be friends a year later? And if you repeated the same scenario with portfolio sizes of 5, 10, 20 and 50 stocks per person, would that change the outcome? Let's see.

```
portfolio <- function(x) {
  stock_data %>%
    slice_sample(n = x, replace = TRUE) %>%
    summarise(
      portfolio_return = mean(return),
      portfolio_size = x
    ) %>%
    bind_rows()
}

set.seed(456)

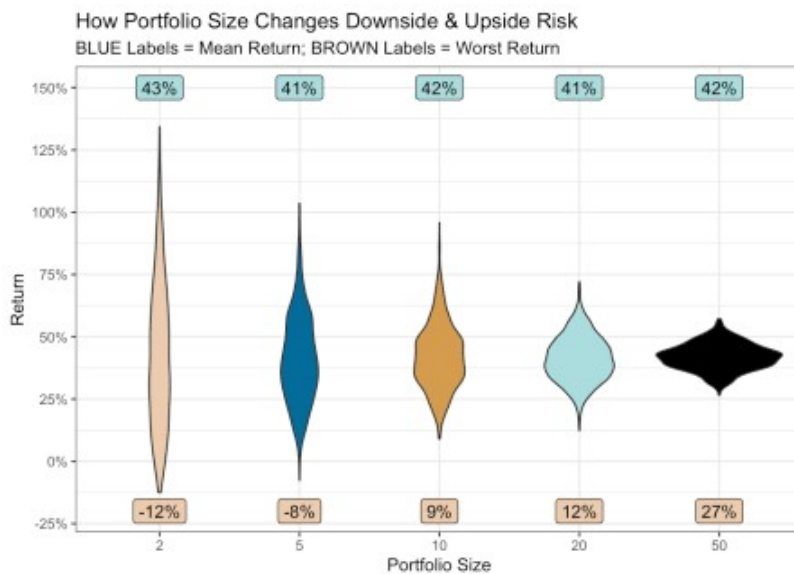
portfolios <-
  map_dfr(c(
    rep(2, 1000),
    rep(5, 1000),
    rep(10, 1000),
    rep(20, 1000),
    rep(50, 1000)
  ), portfolio) %>%
  mutate(portfolio_size = factor(portfolio_size))

mean_returns <- portfolios %>%
  group_by(portfolio_size) %>%
  summarise(
    mean_return = mean(portfolio_return),
    min_return = min(portfolio_return)
  ) %>%
  ungroup()
```

```

portfolios %>%
  ggplot(aes(portfolio_size, portfolio_return, group = portfolio_size)) +
  geom_violin(aes(fill = portfolio_size), show.legend = FALSE) +
  geom_label(aes(portfolio_size, 1.5, label = percent(mean_return, accuracy =
1))),
  data = mean_returns, fill = cols[4],
) +
  geom_label(aes(portfolio_size, -0.2, label = percent(min_return, accuracy =
1))),
  data = mean_returns, fill = cols[1],
) +
  scale_y_continuous(labels = percent_format(), breaks = breaks_extended(9)) +
  scale_fill_manual(values = cols) +
  labs(
    x = "Portfolio Size", y = "Return",
    title = "How Portfolio Size Changes Downside & Upside Risk",
    subtitle = "BLUE Labels = Mean Return; BROWN Labels = Worst Return"
  )

```



So, for all portfolio sizes, the average return across your 1,000 friends is around 42%.

But, when the portfolio size is 2, you may be erased from a few Christmas card lists (or worse). If one of those two stocks has an extreme negative outcome, there's little else in the portfolio to dissipate the effect. As the portfolio size increases, the risk (downside and upside) dramatically reduces.

But is more always better? Well, irrespective of whether your list of promising stocks resulted from desk research or a model, there will be a varying degree of confidence in the 50. A machine learning model, for example, would assign class probabilities to each stock.

So by picking a smaller number, one can select those in which one feels most confident, or which have the highest class probability. And by picking a larger number (ideally across different sectors to further reduce risk) one can weaken the effects of a bad egg or two caused by events no research or model could foresee.

So perhaps the answer is to pick a worst-case scenario one would be prepared to accept. In the plot above, accepting a small chance of only a 12% return (still better than the historical average market return), might provide the "just right" portfolio. A portfolio of a manageable size, focused on your highest-confidence stocks, and with pretty good odds of the desired return.

## R Toolbox

Summarising below the packages and functions used in this post enables me to separately create a [toolbox visualisation](#) summarising the usage of packages and functions across all posts.

Package	Function
base	library[6]; rep[5]; mean[3]; function[2]; return[2]; set.seed[2]; c[1]; chartr[1]; conflicts[1]; cumsum[1]; factor[1]; min[1]; sample[1]; search[1]; sum[1]
dplyr	mutate[5]; filter[4]; if_else[3]; summarise[3]; tibble[3]; group_by[2]; arrange[1]; as_tibble[1]; bind_rows[1]; desc[1]; select[1]; slice_sample[1]; ungroup[1]
ggplot2	aes[5]; geom_label[2]; ggplot[2]; labs[2]; geom_histogram[1]; geom_violin[1]; scale_fill_manual[1]; scale_x_continuous[1]; scale_y_continuous[1]; theme_bw[1]; theme_set[1]
kableExtra	kable[1]
purrr	map[1]; map_dfr[1]; map2_dfr[1]; possibly[1]; set_names[1]
readr	cols[1]; read_lines[1]
rebus	literal[4]; lookahead[3]; whole_word[2]; ALPHA[1]; lookbehind[1]; one_or_more[1]; or[1]
scales	percent[2]; percent_format[2]; breaks_extended[1]
stringr	str_detect[3]; str_c[2]; str_remove[2]; str_count[1]; str_remove_all[1]
tibble	enframe[1]
tidyr	tibble[3]; as_tibble[1]; unnest[1]
truncnorm	rtruncnorm[1]
wesanderson	wes_palette[1]