There is no doubt that, every now and then, one ought to celebrate life. This usually involves people coming together, talking, laughing, dancing, singing, shouting; simply put, it means throwing a party. With temperatures rising, summer offers all the more incentive to organize such a joyous event. Blinded by the light, it is easy to forget that we are, unfortunately, still in a pandemic. But should that really deter us?

Walking around central Amsterdam after sunset, it is easy to notice that not everybody holds back. Even if my Dutch was better, it would likely still be difficult to convince groups of twenty-somethings of their potential folly. Surely, they say, it is exceedingly unlikely that this little party of ours results in any virus transmission?

Government retorts by shifting perspective: while the chances of virus spreading at any one party may indeed be small, this does not licence throwing it. Otherwise many parties would mushroom, considerably increasing the chances of virus spread. Indeed, government stresses, this is why such parties remain *illegal*.

But while *if-everybody-did-what-you-did* type of arguments score high with parents, they usually do no score high with their children. So instead, in this blog post, we ask the question from an individual’s perspective: what are the chances of getting the virus after attending this or that party? And what factors make this more or less likely?

**Risky business?**

To get us started, let’s define the *risk of a party* as the probability that somebody who is infected with the novel coronavirus and can spread it attends the gathering. The two major factors influencing this probability are the size of the party, that is, the number of people attending; and the prevalence of infected people in the relevant population. As we will see, the latter quantity is difficult to estimate. The probability of actually getting infected by a person who has the coronavirus depends further on a number of factors; we will discuss those in a later section.

I will base the following calculations on data from the Netherlands, specifically from Amsterdam. You can exchange these numbers with numbers from your country and city of choice. RIVM — the Netherland’s institute for public health — reports new cases across regions. Between July $8^{\text{th}}$ and July $21^{\text{st}}$, there have been $20.4$ reported new cases per $100,000$ residents in the [region of Amsterdam](https://www.rivm.nl/en/novel-coronavirus-covid-19/current-information), yielding a total of $178$ cases. In the next section, we discuss the difficulties in estimating the prevalence of infections.

**Estimating the prevalence of infections**

While $178$ is the *reported* number of new cases, it is not the *true* number of new cases.[1](https://fabiandablander.com/r/Corona-Party.html#fn:1) How do we arrive at that?

First, note that the number of *new* cases is not the number of *current* cases. We care about the latter. More specifically, we care about the number of currently *infectious* cases. This is different from the number of currently *infected* cases.

These are complications, but we need to keep it simple. Currently, visitors from outside Europe must show a negative COVID-19 test or need to self-isolate for $14$ days upon arrival in most European countries. Let’s take these $14$ days for simplicity, and assume conservatively that this is the time one is infectious upon getting infected. Thus, we simply take the reported number of *new infected* cases in the last two weeks as the reported number of *currently infectious* cases.

With that out of the way, a key question remains: how do we get from the *reported* number of infections to the *true* number of infections? One can estimate the true number of infections using models, or by empirically estimating the seroprevalence in the population, that is, the proportion of people who have developed antibodies.

Using the first approach, Flaxman et al. (2020) estimate that the total percentage of the population that has been infected — the *attack rate* — across $11$ European countries as of May $4^{\text{th}}$. The Netherlands was, unfortunately, not included in these estimates. To get some intuition for the calculation, and to see how countries that have fared better or worse compare to the Netherlands, we turn to Germany, Spain, and the UK. For these three countries the estimated true attack rates were $0.85\%$, $5.50\%$, and $5.10\%$, respectively. Given the population of these countries and the cumulative number of reported infections, we can compute the *reported* attack rate. Relating this to the estimate of the *true* attack rate gives us an indication of the extent that the reports undercount the actual infections; the code below calculates this for the three countries.

library('dplyr')

library('COVID19')

get\_undercount <- function(country, attack\_rate) {

# Flaxman et al. (2020) estimate the attack rate as of 4th of May

dat <- covid19(country = country, end = '2020-05-04', verbose = FALSE)

dat %>%

group\_by(id) %>%

summarize(

population = head(population, 1),

total\_cases = tail(confirmed, 1)

) %>%

mutate(

attack\_rate = attack\_rate,

reported\_attack\_rate = 100 \* total\_cases / population,

undercount\_factor = attack\_rate / reported\_attack\_rate

)

}

get\_undercount(c('Germany', 'Spain', 'United Kingdom'), c(0.85, 5.5, 5.10))

## # A tibble: 3 x 6

## id population total\_cases attack\_rate reported\_attack\_rate undercount\_factor

##

## 1 DEU 82905782 165120 0.85 0.199 4.27

## 2 ESP 46796540 218011 5.5 0.466 11.8

## 3 GBR 66460344 191843 5.1 0.289 17.7

The table above shows that cases were undercounted by a factor of about $4$ in Germany, $12$ in Spain, and $18$ in the UK. The Netherlands undercounted cases by a factor of about $10$ in April (Luc Coffeng, personal communication). The attack rate estimate for Spain is confirmed by a recent seroprevalence study, which finds a similarly low overall proportion of people who have developed antibodies (around $5\%$, with substantial geographical variability) in the period between April $27^{\text{th}}$ and May $11^{\text{th}}$ (Pollán et al. 2020). In another seroprevalence study, Havers et al. 2020 find that between March $23^{\text{rd}}$ and May $12^{\text{nd}}$, reported cases from several areas in the United States undercounted true infections by a factor between $6$ and $24$.

Currently, the pandemic is not as severe in Europe as it was back when the above studies were conducted. Most importantly, the testing capacity has been ramped up. For example, while the proportion of positive tests in Germany, Spain, the UK, and the Netherlands were $2.40\%$, $2.60\%$, $7.10\%$, and $9.40\%$ on the $4^{\text{th}}$ of May (the end date used in the Flaxman et al. 2020 study), they are $0.50\%$, $1.40\%$, $0.60\%$, and $0.60\%$ respectively, using most recent data from [here](https://ourworldindata.org/coronavirus-testing) at the time of writing. Thus, these countries are tracking the epidemic much more closely, which in turn implies that the factor by which they undercount the true cases is likely lower than it was previously.

At the same time, cases are rising again, and RIVM estimates the effective reproductive number to be $R\_t = 1.29$. Let’s assume therefore that the true number of *infectious* cases is $5$ times higher then the number of reported *infected* cases. This includes asymptomatic carriers or those that are pre-symptomatic but still can spread the virus. We assume the estimated *true* number of infectious cases to therefore be $5 \times 20.4 = 102$ per $100,000$ residents in Amsterdam. We will test how robust our results are against this particular choice later; in the next section, we estimate the risk of a party.

**Estimating the risk of a party**

What are the chances that a person who attends your party has the coronavirus and is infectious? To calculate this, we assume that party guests form an independent random sample from the population. We will discuss the implications of this crude assumption later; but for now, it allows us to estimate the desired probability in a straightforward manner.

We estimated that Amsterdam had $5 \times 20.4 = 102$ true infectious cases per $100,000$ inhabitants. Assuming that the probability of infection is the same for all citizens (more on this later), this results in $102 / 100,000 = 0.00102$, which gives a $0.102\%$ or $1$ in $980$ chance that a *single* party guest has the virus and can spread it.

A party with just one guest would be — *intimate*. So let’s invite a few others. What are the chances that *at least one* of them can spread the virus? We compute this by first computing the complement, that is, the probability that *no* party guest is infectious.

The chance that any one person from Amsterdam is not infectious is $1 – 0.00102 = 0.9990$, or $99.90\%$. With our assumption of guests being an independent random sample from the population, the probability that none of the $n$ guests can spread the virus is $0.9990^n$.

The functions below compute the probability that no party guests can spread the virus, as well as the probability that *at least one* of the guests is infectious; the latter probability is simply the complement of the former.

# Probability that no guest is infectious

prob\_virus\_free <- function(n, true\_relative\_cases = 20.4 \* 5) {

prob\_virus <- true\_relative\_cases / 100000

(1 - prob\_virus)^n

}

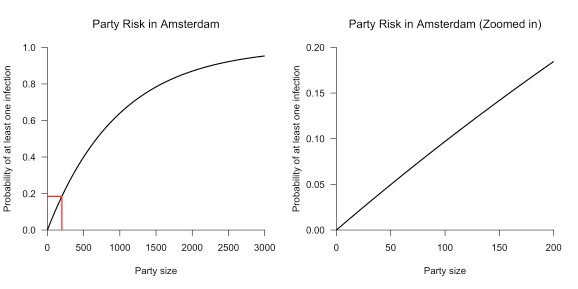
# Probability that at least one guest is infectious

party\_risk <- function(n, true\_relative\_cases = 20.4 \* 5) {

1 - prob\_virus\_free(n, true\_relative\_cases)

}

The figure below shows the party risk in Amsterdam as a function of the party size $n$.



The left panel shows how the party risk — the probability that at least one infectious person is the party — increases with $n$. In particular, to have near certainty that at least one infectious person shows up requires a very large party. The right panel zooms in on reasonably party sizes. Most parties that are thrown indoors probably do not exceed $100$ attendants, depending on how rich and reckless the host is.

The estimate of the party risk based on this simple calculation are somewhat sobering: there is a $2.02\%$, a $4.97\%$, and a $14.19\%$ chance that at least one guest can spread the coronavirus for parties of size $20$, $50$, and $150$, respectively. We have of course made a number of simplifying assumptions to arrive at these estimates, and we will critically discuss them in a later section. We will also talk about the factors that influence the chances of actually getting infected when an infectious guest shows up.

In the terms of our calculations, their model applies a correction factor of $37.2 / 20.4 = 1.824$. RIVM is therefore slightly more optimistic than I am; for parties of size $20$, $50$, and $150$, their estimates of the probability that at least one guest is infectious — assuming guests are a random sample from the population — are $0.74\%$, $1.84\%$, and $5.43\%$, respectively.

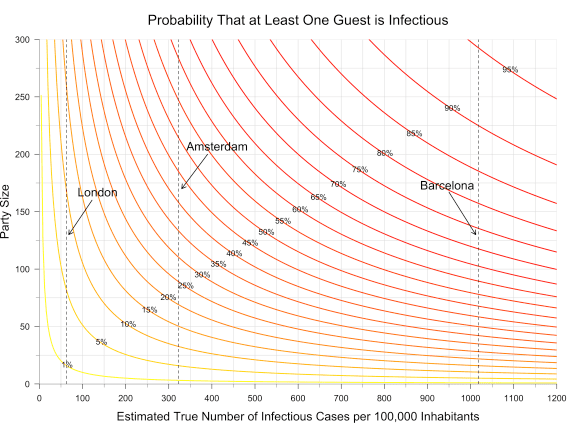
In any event, what is important to note is that the estimates change with the correction factor. In the next section, we assess this relationship moree systematically.

**Sensitivity analysis**

We have assumed that the reported number of infected cases undercounts the true number of infectious cases by a factor of $5$. In particular, we used the reported cases of $20.4$ per $100,000$ and applied a correction factor of $5$. But what if, in a week from now, the reported cases are $30$ per $100,000$? In the following, we visualize the party risk as a function of the estimated *true* infectious cases per $100,000$ inhabitants (see also Lachmann & Fox, 2020).

Between July $8^{\text{th}}$ and July $21^{\text{st}}$, $20.4$ cases per $100,000$ inhabitants were reported in Amsterdam. A correction factor of $5$ would bolster this to $102$ cases, a correction factor of $10$ to $204$ cases, and so on; thus, one can backcalculate the correction factor from the estimated true number of cases.

The figure below visualizes the probability that at least one party guest has the coronavirus and can spread it as a function of the estimated *true* number of infectious cases per $100,000$ inhabitants and the size of the party.



Let’s take a moment to unpack this figure. Each coloured line represents a combination of true number of infectious cases and party size that yields the same party risk. For example, attending a party of size $20$ when the true number of infectious cases per $100,000$ inhabitants is $50$ yields a party risk of $1\%$, but so would, roughly, attending a party of size $10$ when the true relative number of infectious cases is $100$. Thus, there is a trade-off between the size of the party and the true number of infectious cases.

You can get a quick overview of the party risk for different party sizes by checking when the gray solid lines *verticallly* cross the coloured lines. Similarly, you can get a rough understanding for the party risk for different relative numbers of true infectious by checking when the gray and coloured lines cross *horizontally*. The dotted vertical line in the figure gives our previous estimate of the true number of cases.

The figure allows you to estimate the party risk wherever you live; just look up the local number of new cases in the last two weeks and, making the assumptions we have made so far, the plot above gives you the probability that at least one party guest will turn up infectious with the coronavirus. The assumptions we have made are very simplistic, and indeed, if you have a more elaborate way of estimating the number of currently infectious cases, then you can use that number combined with the figure to estimate the party risk.

While we have computed the party risk for a single party, this risk naturally increases when you attend multiple ones. Suppose you have been invited to parties of size $20$, $35$, and $50$ which will take place in the next month. Let’s for simplicity assume that all guests are different each time. Let’s further assume that the number of infectious cases stays constant over the next month; with a current estimate of $R\_t = 1.29$, this seems unlikely. Together, these assumptions allow us to calculate the *total* party risk as the party risk of attending a single party of size $105$, which gives $10.16\%$. It seems that, in this case, fortune does not favour the bold.

**Assumptions**

The analysis above is a very rough *back-of-the-envelope* calculation. We have made a number of crucial assumptions to arrive at some numbers. That’s useful as a first approximation; now we have at least some intuition for the problem, and we can critically discuss the assumptions we made. Most importantly, do these assumptions lead to *overestimates* or *underestimates* if the party risk?

**Independence**

First, and most critically, we have assumed that party guests are a *randomly and independently* drawn from the population. It is this assumption that allowed us to compute the joint probability that none of the party guests have the virus by multiplying the probabilities of any individual being virus-free. If you have ever been to a party, you know that this is not true: instead, a considerable number of party guests usually know each other, and it is safe to say that they are similar on a range of socio-demographic variables such as age and occupation.[3](https://fabiandablander.com/r/Corona-Party.html#fn:3)

This means we are sampling not from the whole population, as our simple calculation assumes, but from some particular subpopulation that is well connected. Since the party guests likely share social circles or even households, the *effective* party size — in terms of being relevant for virus transmission — is smaller than the *actual* party size; this is because these individuals share the same risks. A party with $20$ married couples seems safer than a party with $40$ singles. This would suggest that we overestimate the risk of a party.

**Uniform infection probability**

At the same time, however, our calculations assume that the risk of getting the coronavirus is evenly spread across the population. We used this fact when estimating the probability that any one person has the coronavirus as the total number of cases divided by the population size.

The probability of infection is not, howevever, evenly distributed. For example, Pollán et al. (2020) report a seroprevalence of people aged $65$ or more of about $6\%$, while people aged between $20$ and $34$ showed a seroprevalence of $4.4\%$ between April $27^{\text{th}}$ to May $11^{\text{th}}$. These days, however, there is a substantial rise in young people who become infected, in the [United States](https://www.nytimes.com/2020/06/25/us/coronavirus-cases-young-people.html) and likely also in Europe. Because young people are [less likely to develop symptoms](https://www.axios.com/coronavirus-young-people-spread-5a0cd9e0-1b25-4c42-9ef9-da9d9ebce367.html), the virus can spread largely undetected.

Moreover, it seems to me that people who would join a party are in general more *adventurous*. This would increase the chances of an infectious person attending a party; thus, our calculation above may in fact underestimate the party risk.

At the same time, one would hope that people who show symptoms avoid parties. If all guests do this, then only pre-symptomatic or asymptomatic spread can occur, which would reduce the party risk by a half up to two thirds. On the flip side, people who show symptoms might get tested for COVID-19 and, upon receiving a negative test, consider it safe to attend a party. This might be foolish, however; recent estimates suggest that tests miss about $20\%$ infections for people who show symptoms (Kucirka et al, [2020](https://www.acpjournals.org/doi/10.7326/M20-1495); see also [here](https://www.advisory.com/daily-briefing/2020/07/06/negative-covid)). For people without symptoms, the test performs [even worse](https://www.theatlantic.com/science/archive/2020/06/how-negative-covid-19-test-can-mislead/613246/).

For parties taking place in summer, it is not unlikely that many guests engaged in holiday travel in the days or weeks before the date of the party. Since travel increases the chances of infection, this would further increase the chances that at least one party guest has contracted the coronavirus.

**Estimating true infections**

We have assumed that the number of *new* cases in the last two weeks equals the number of currently *infectious* cases. This is certainly an approximation. Ideally, we would have a geographically explicit model which, at any point in time and space, provides an estimate of the number of infectious cases. To my knowledge, we are currently lacking such a model.

Note that, if the $178$ people who tested positive all self-isolate or, worse, end up in hospital, this clearly curbs virus spread compared to when they would continue to roam around. The former seems more likely. Moreover, these reported cases are likely not independent either, with outbreaks usually being localized. Similar to the fact that party guests know each other, the fact that reported cases cluster would lead us to overestimate the extent of virus spread.

At the same time, in the Netherlands only those that show symptoms can get tested. Since about a third to a half are asymptomatic or pre-symptomatic in the sense that they spread the virus considerably before symptom onset, the reported number of cases likely gives an undercount of infectious people.

All these complications can be summarized, roughly, in the correction factor, which gives the extent to which we believe that the reported number of cases deviates from the true number of infectious cases. We have first focused on a factor of $5$, but then assessed the robustness of our results in a sensitivity analysis. For example, for a party size of $50$, the chances that at least one guest is infectious is $1.84\%$ for a correction factor of $1.824$ (corresponding to the official RIVM estimate), $4.97\%$ for a factor of $5$, and $18.49\%$ for a factor of $20$. You can play around with these numbers yourself. Observe how they make you feel. Personally, given what we said above about the infection probability for young and adventurous people, I am inclined to err on the side of caution.

**Estimating the probability of infection**

We have the defined the party risk as the probability that at least one party guest has the coronavirus and is infectious. If this person does not spread the virus to other guests, no harm is done.

This is exceedingly unlikely, however. The probability of getting infected is a function of the time one is exposed to the virus, and the amount of virus one is exposed to. [Estimates suggest](https://www.erinbromage.com/post/the-risks-know-them-avoid-them) that about $1,000$ SARS-CoV-2 infectious virus particles suffice for an infection. With breathing, about $20$ viral particles diffuse into the environment per minute; this increases to $200$ for speaking; coughing or sneezing can release $200,000,000$ (!) virus particles. These do not all fall to the ground, but instead can remain suspended in the air and fill the whole room; thus, physical distancing alone might [not be enough indoors](https://www.nytimes.com/2020/07/06/health/coronavirus-airborne-aerosols.html) (the extent of airborne transmission remains debated, however; see for example Klompas, Baker, & Rhee, [2020](https://jamanetwork.com/journals/jama/fullarticle/2768396)). It seems reasonable to assume that, when party guests are crowded in a room for a number of hours, many of them stand a good chance of getting infected if any one of the guests is infectious. [Masks would help](https://www.erinbromage.com/post/what-s-the-deal-with-masks), of course; but how would I sip my Negroni, wearing one?

It is different outdoors. A Japanese study found that virus transmission inside was about $19$ times more likely than outside (Nishiura et al. [2020](https://www.medrxiv.org/content/10.1101/2020.02.28.20029272v2)). Analyzing $318$ outbreaks in China between January $4^{\text{th}}$ and February $11^{\text{th}}$, Quian et al. ([2020](https://www.medrxiv.org/content/10.1101/2020.04.04.20053058v1)) found that only a single one occurred outdoors. This suggests that parties outdoors should be much safer than parties indoors. Yet outdoor parties feature elements unlike other outdoor events; for example, there are areas — such as bars or [public toilets](https://www.nytimes.com/2020/06/24/style/coronavirus-public-bathrooms.html) — which could become spots for virus transmission. Our simple calculations suggest, with a correction factor of $5$, that the probability that at least one person out of $150$ has the coronavirus is $14.19\%$. While, in contrast to an indoor setting, the infected person is unlikely to infect the majority of the other guests, it seems likely that at least some guests will get the virus.

**Conclusion: To party or not to party?**

If I do not care whether I get wet or not, I will never carry an umbrella, regardless of the chances of rain. Similarly, my decision to throw (or attend) a party requires not only an estimate of how likely it is that the virus spreads at the gathering; it also requires an assessment of how much I actually care.

As argued above, it is almost certain that the virus spreads to other guests if one guest arrives infected. Noting that all guests are young, one might be tempted to argue that the cost of virus spread is low. In fact, people who party might even be helping — *heroically* — to build [herd immunity](https://fabiandablander.com/r/Covid-Exit.html)!

This reasoning is foolish on two grounds. First, while the proportion of infected people who die is very small for young people — Salje et al. ([2020](https://science.sciencemag.org/content/369/6500/208)) estimate it to be $0.0045\%$ for people in their twenties and $0.015\%$ for people in their thirties — the picture about the non-lethal, long-term effects of the novel coronavirus is only slowly becoming clear. For some people, recovery can be [lengthy](https://www.theatlantic.com/health/archive/2020/06/covid-19-coronavirus-longterm-symptoms-months/612679/) — much longer than the two weeks we previously believed it would take. Known as “mild” cases, they [might not be so mild after all](https://www.theguardian.com/commentisfree/2020/jul/06/coronavirus-covid-19-mild-symptoms-who). Moreover, the potential [strange neurological effects](https://www.bbc.com/future/article/20200622-the-long-term-effects-of-covid-19-infection) of a coronavirus infection are becoming increasingly apparent. All told, party animals, even those guarded by their youth, might not shake it off so easily.

Suppose that, even after carefully considering the potential health dangers, one is still willing to take the chances. After all, it would be a *really* good party, and we young people usually eat our veggies — especially in Amsterdam. The trouble with infectious diseases, though, is that they travel: while you might be happy to take a chance, you and the majority of party guests will probably not self-quarantine after the event, right? If infections occur at the party, the virus is thus likely to subsequently spread to other, more vulnerable parts of the population.

So while you might remain unharmed after attending a party, others might not. Take the story of [Bob from Chicago](https://www.erinbromage.com/post/the-risks-know-them-avoid-them), summarizing an actual infection chain reported by Ghinai et al. ([2020](https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e1.htm)):

“Bob was infected but didn’t know. Bob shared a takeout meal, served from common serving dishes, with $2$ family members. The dinner lasted $3$ hours. The next day, Bob attended a funeral, hugging family members and others in attendance to express condolences. Within $4$ days, both family members who shared the meal are sick. A third family member, who hugged Bob at the funeral became sick. But Bob wasn’t done. Bob attended a birthday party with $9$ other people. They hugged and shared food at the $3$ hour party. Seven of those people became ill.

But Bob’s transmission chain wasn’t done. Three of the people Bob infected at the birthday went to church, where they sang, passed the tithing dish etc. Members of that church became sick. In all, Bob was directly responsible for infecting $16$ people between the ages of $5$ and $86$. Three of those $16$ died.”

These events took place before much of the current corona measures were put in place, but the punchline remains: parties are a matter of public, not only individual health. Don’t be like Bob.

**Post Scriptum**

The code below reproduces the first figure in the main text.

plot\_risk <- function(ns, true\_relative\_cases = 20.4 \* 5, ...) {

plot(

ns, 1 - prob\_virus\_free(ns, true\_relative\_cases = true\_relative\_cases),

type = 'l', axes = FALSE,

xlab = 'Party size', ylab = 'Probability of at least one infection',

xaxs = 'i', yaxs = 'i', ...

)

axis(1)

axis(2, las = 2)

}

par(mfrow = c(1, 2))

n1 <- 3000

n2 <- 200

y2 <- 1 - prob\_virus\_free(n2)

plot\_risk(

seq(0, n1), lwd = 2, main = 'Party Risk in Amsterdam',

xlim = c(0, n1), ylim = c(0, 1), font.main = 1.5

)

lines(c(n2, n2), c(0, y2), col = 'red', lwd = 2)

lines(c(0, n2), c(y2, y2), col = 'red', lwd = 2)

plot\_risk(

seq(0, n2), lwd = 2, main = 'Party Risk in Amsterdam (Zoomed in)',

ylim = c(0, 0.20), xlim = c(0, n2), font.main = 1.5

)

The code below reproduces the second figure in the main text.

library('RColorBrewer')

# Calculates the party size that results in 'prob\_virus\_free'

# for a given 'true\_relative\_cases'

get\_party\_size <- function(prob\_virus\_free, true\_relative\_cases) {

log(prob\_virus\_free) / log(1 - true\_relative\_cases / 100000)

}

plot\_total\_risk <- function(party\_sizes, true\_relative\_cases, ...) {

plot(

true\_relative\_cases, ns, type = 'n', xaxs = 'i',

yaxs = 'i', axes = FALSE, ylab = 'Party Size',

xlab = 'True Number of Infectious Cases per 100,000 Inhabitants', ...

)

ticks <- seq(0, 300, 25)

axis(1, at = ticks)

axis(2, at = ticks, las = 2)

abline(h = ticks, col = 'gray86')

abline(v = ticks, col = 'gray86')

probs\_virus <- seq(0.01, 0.99, 0.01)

party\_sizes <- sapply(1 - probs\_virus, get\_party\_size, true\_relative\_cases)

cols <- rev(heat.colors(50))

cols <- colorRampPalette(cols, bias = 2)(99)[-1]

ix <- c(seq(1, 55, 1), seq(60, 95, 5))

show\_text <- c(seq(1, 10, 1), seq(15, 55, 5))

for (i in ix) {

y <- party\_sizes[, i]

lines(true\_relative\_cases, y, col = cols[i], lwd = 2.5)

j <- which.min(true\_relative\_cases[-1] + y[-1])

if (i %in% show\_text) {

text(

true\_relative\_cases[j] + 1, y[j] + 1, paste0(probs\_virus[i] \* 100, '%'),

cex = 0.80

)

}

}

}

ns <- seq(0, 300)

true\_relative\_cases <- seq(0, 300, length.out = length(ns))

plot\_total\_risk(

ns, true\_relative\_cases,

main = 'Probability That at Least One Guest is Infectious',

font.main = 1, cex.main = 1.5

)

lines(c(20.4 \* 5, 20.4 \* 5), c(0, 300), lty = 2)