Does position of a motif influence functionality.

1. Motif: (?=[WYF].{0,4}[DE]|[DE].{0,4}[WYF])

For this analysis I got the subset of random library that has this motif.

I separated sequences into 20 subsets based on length and found live\_percent of the motif vs the distance from the N-terminus.

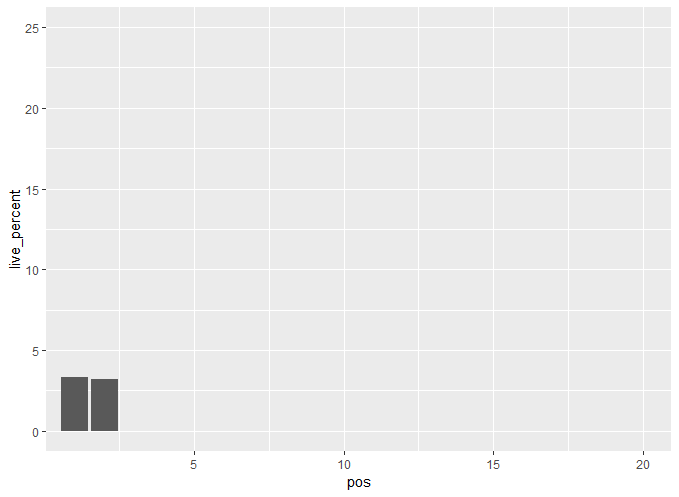
S P Y A P Y D R V

9 8 7 6 5 4 3 2 1

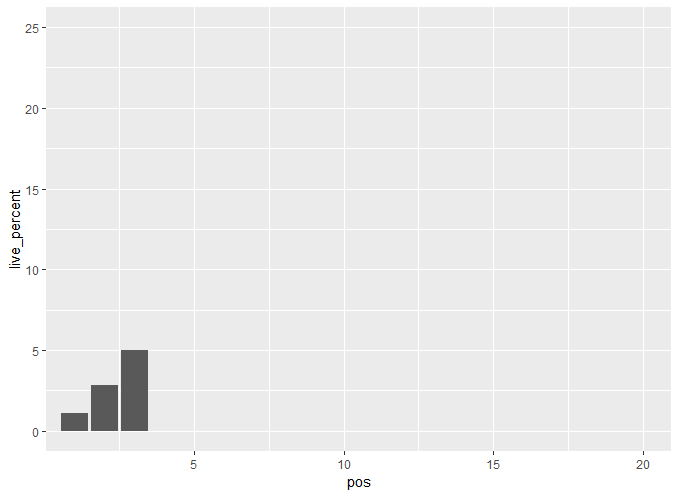
Length 2 – 0 live, 29 die

2 1

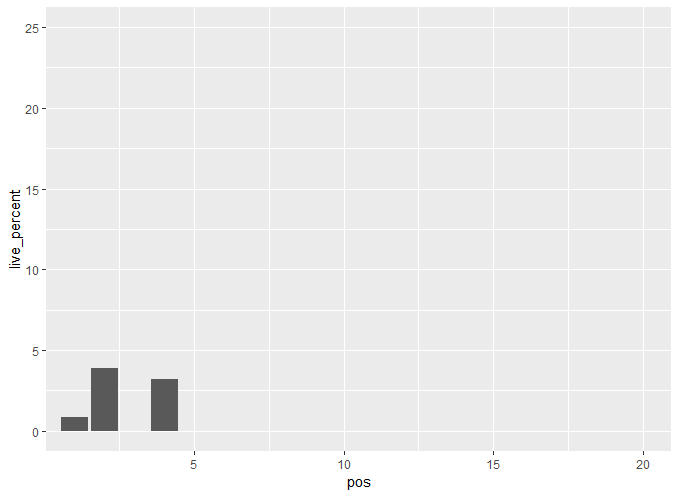
Length 3 – 2 live, 84 die



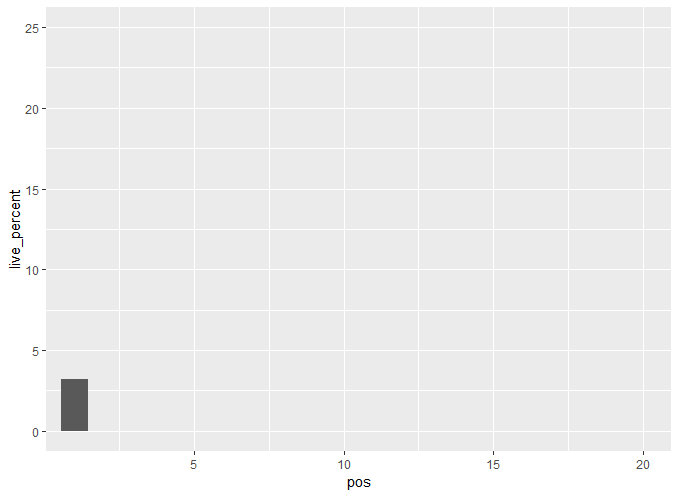
Length 4 – 3 live, 161 die



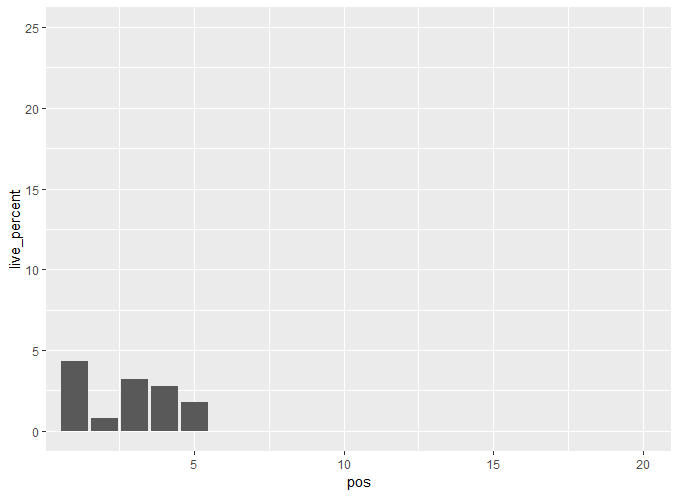
Length 5 – 5 live, 233 die



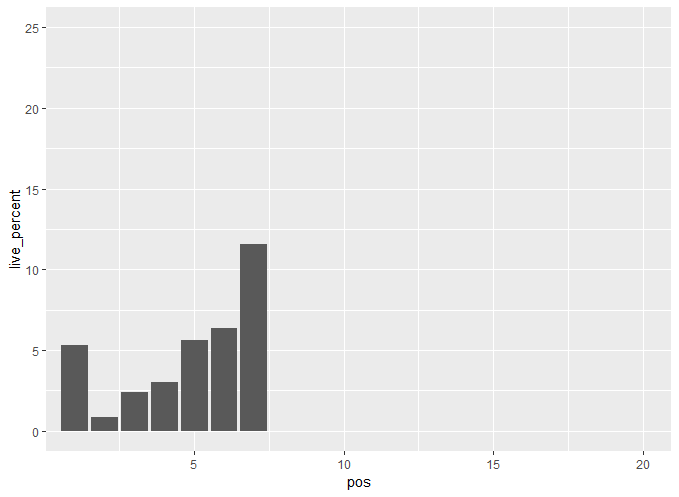
Length 6 – 4 live, 309 die



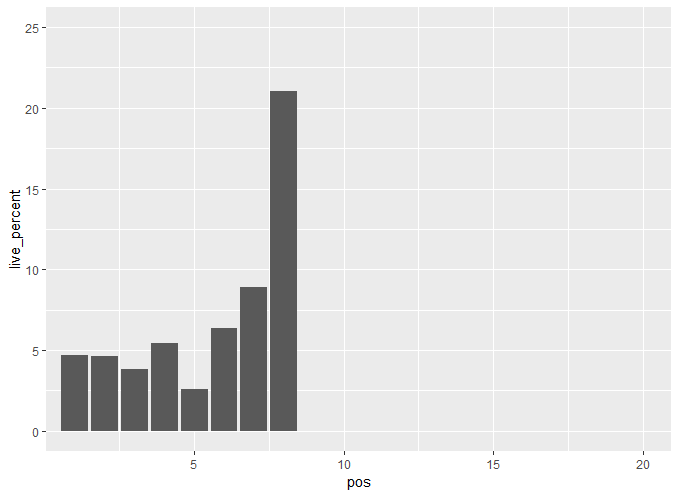
Length 7 – 9 live, 367 die



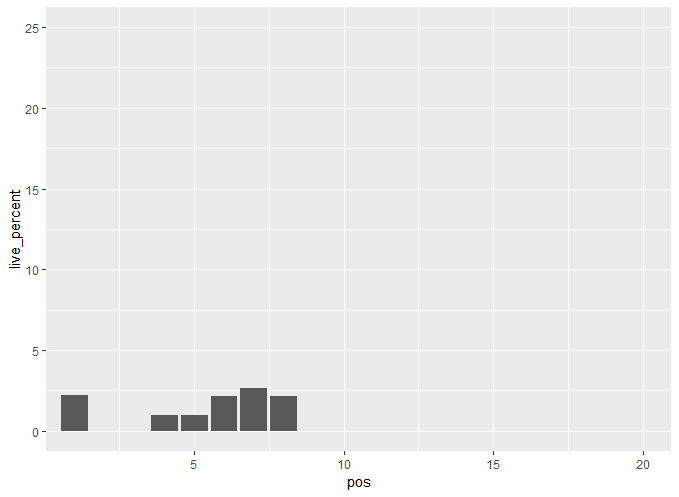
Length 8 – 11 live, 404 die



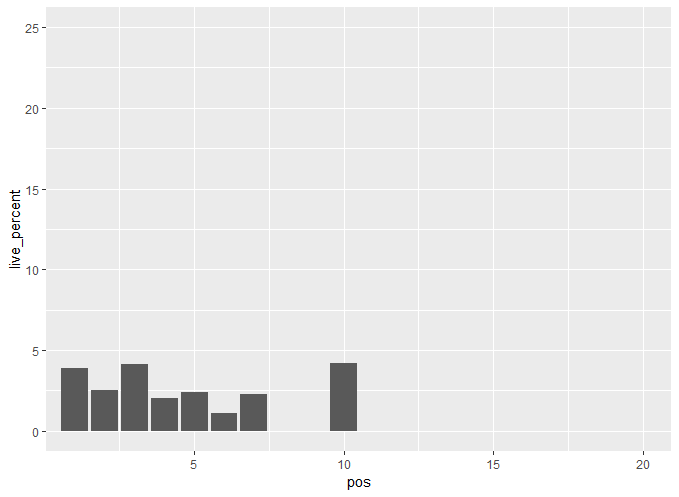
Length 9 – 11 live, 441 die



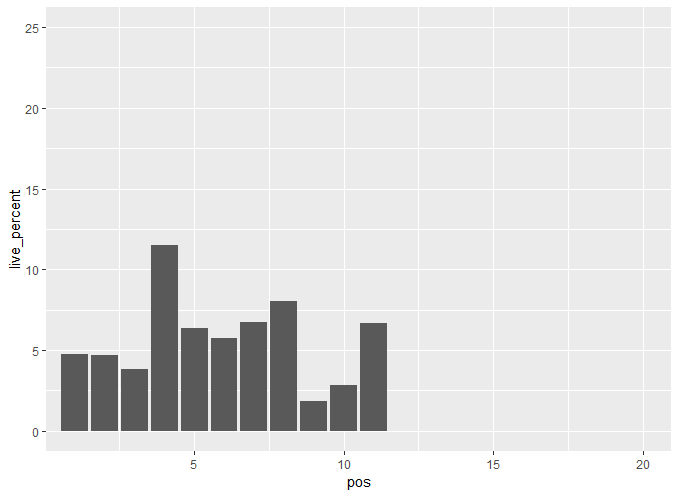
Length 10 – 5 live, 509 die



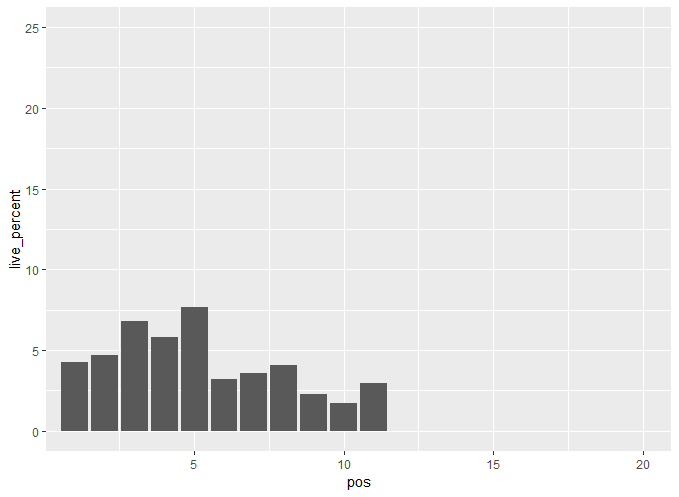
Length 11 – 8 live, 508 die



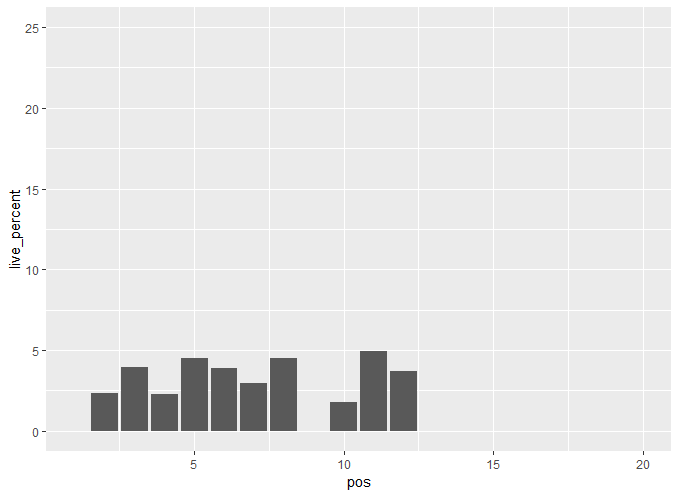
Length 12 – 23 live, 492 die



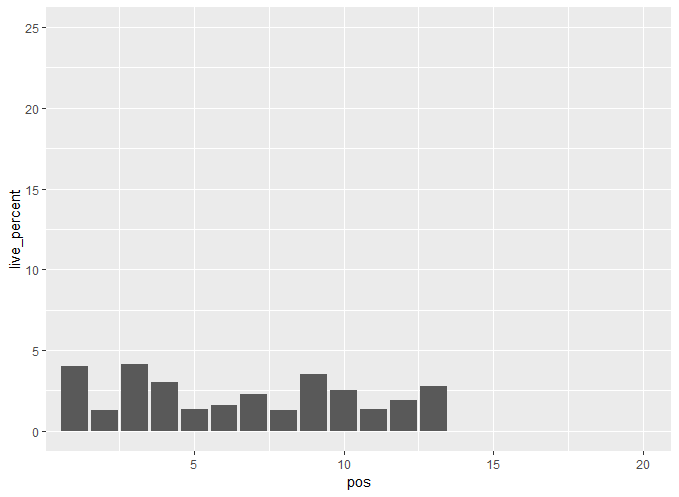
Length 13 – 19 live, 549 die



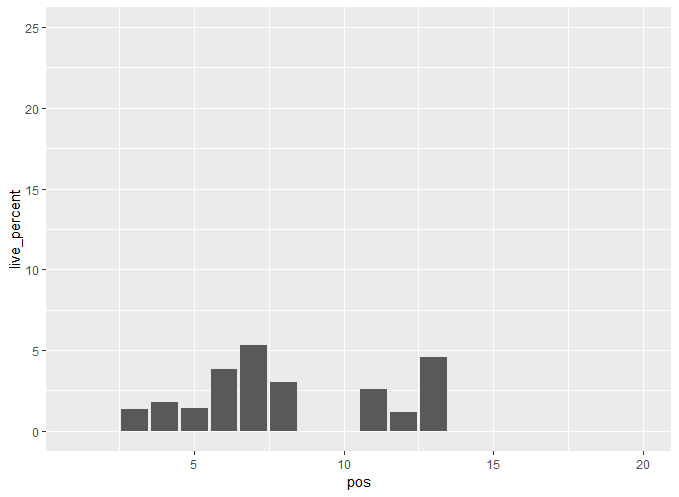
Length 14 – 10 live, 529 die



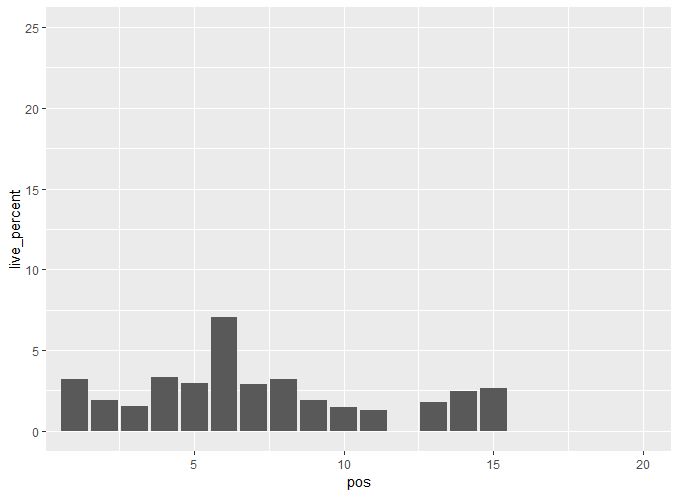
Length 15 – 6 live, 538 die



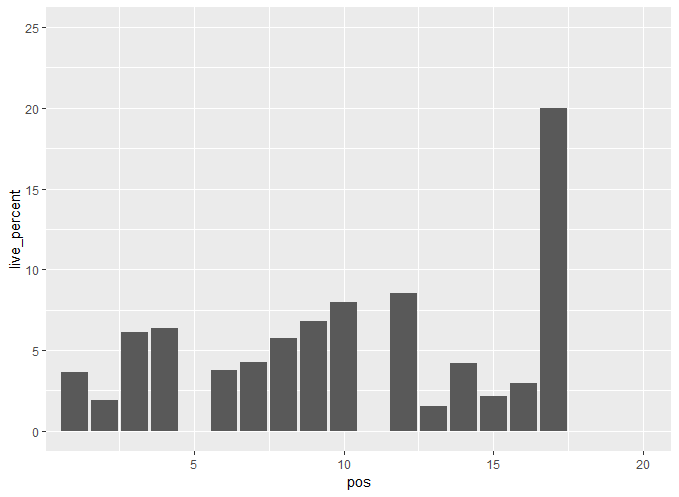
Length 16 – 9 live, 537 die



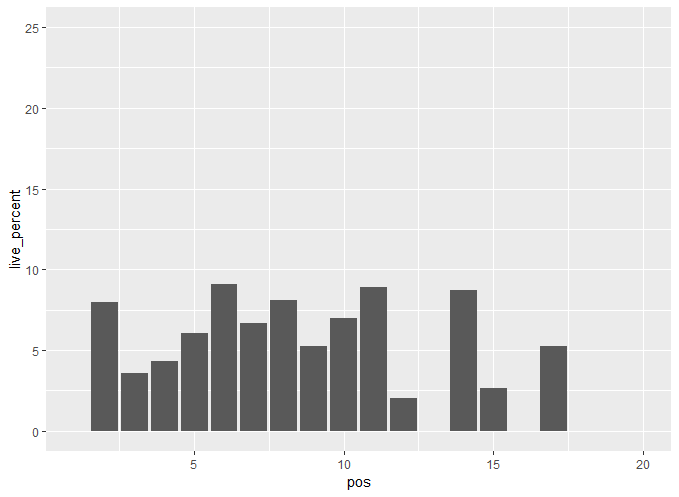
Length 17 – 10 live, 540 die



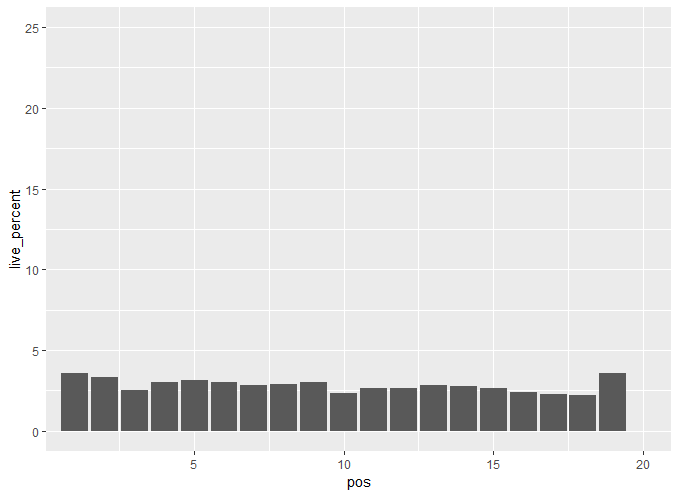
Length 18 – 13 live, 474 die



Length 19 – 13 live, 403 die



Length 20 – 326 live, 12433 die

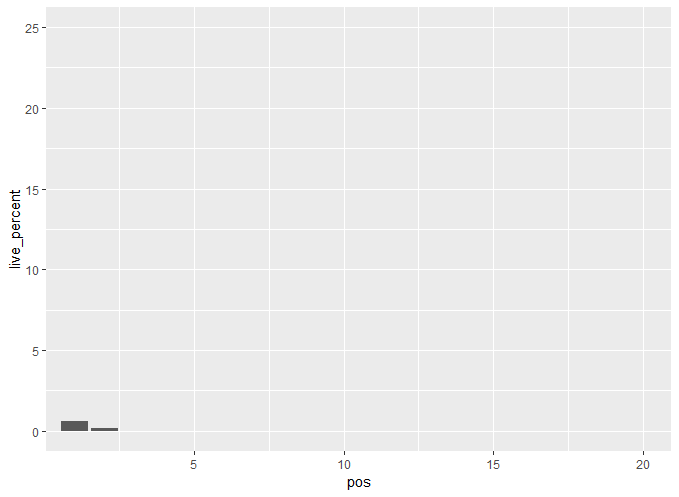


Conclusion: Length 20 subset shows no clear connection between position and functionality. Looking at the smaller subsets I don’t see a trend that is consistently shared. Since length 20 has the most data to work and it shows no clear connection, I would say position is not too important.

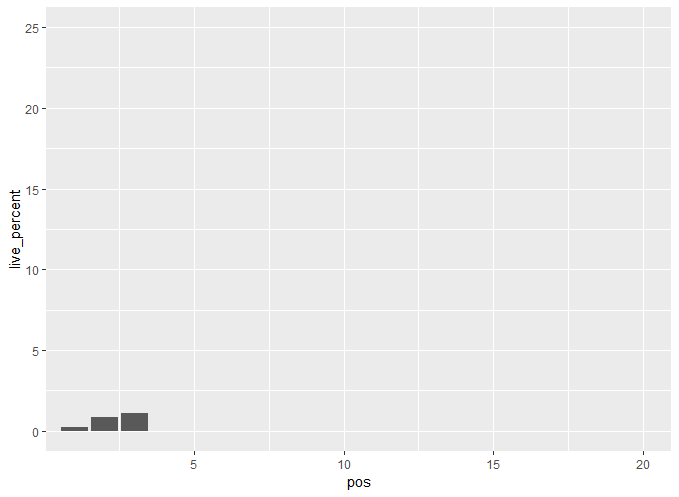
1. Motif: [RHK]

Working under the assumption basic amino acids are detrimental to functionality is there a position at which they are more detrimental

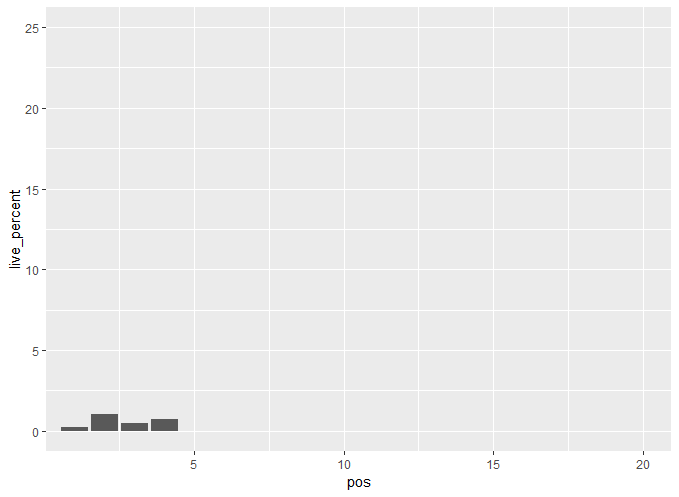
Length 2 – 4 live, 906 die



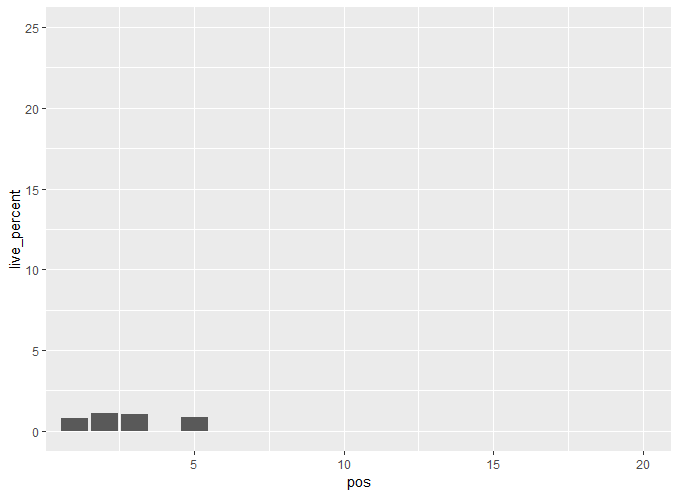
Length 3 – 9 live, 1156 die



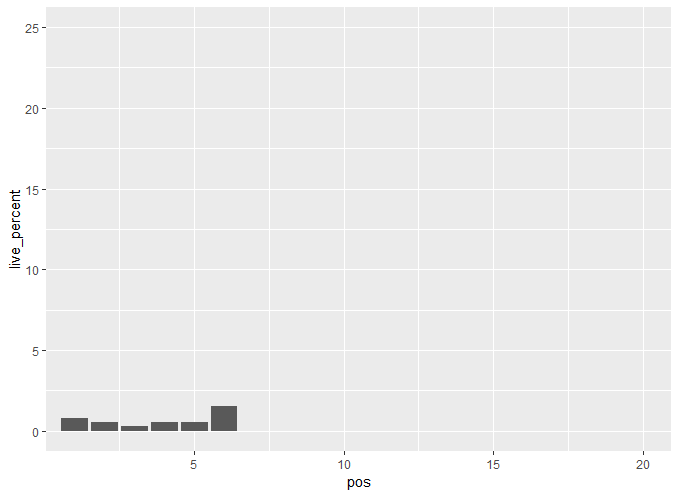
Length 4 – 9 live, 1258 die



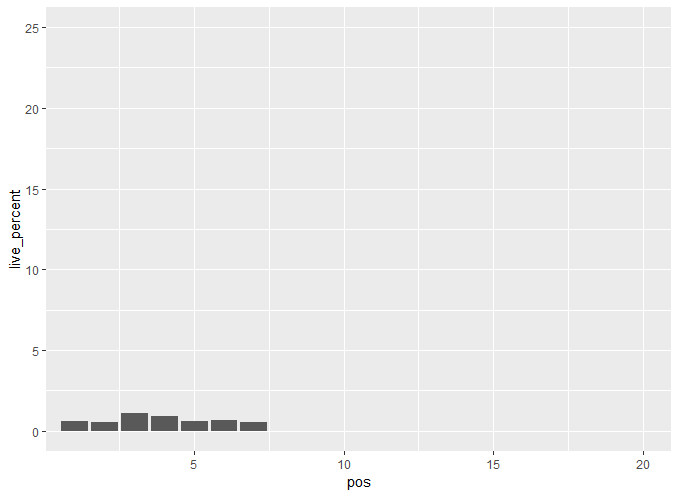
Length 5 – 11 live, 1363 die



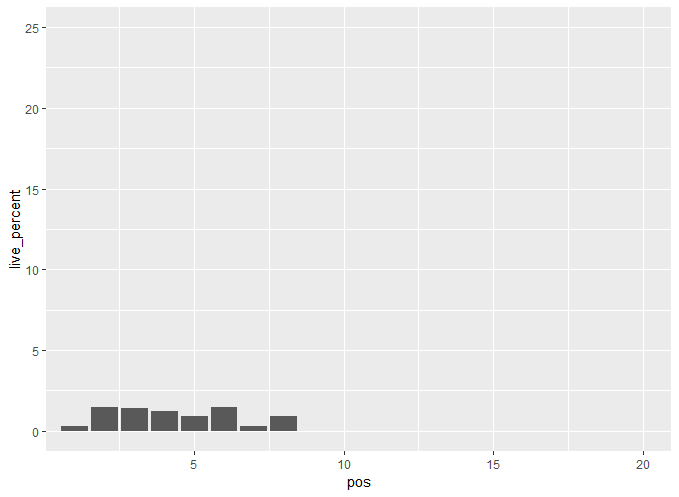
Length 6 – 11 live, 1489 die



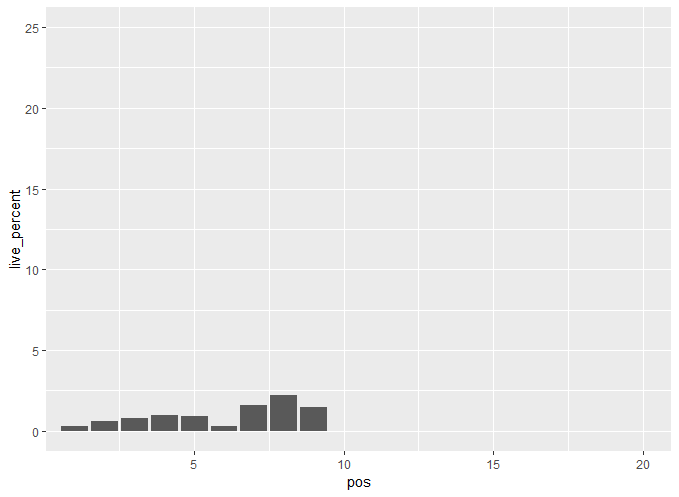
Length 7 – 12 live, 1491 die



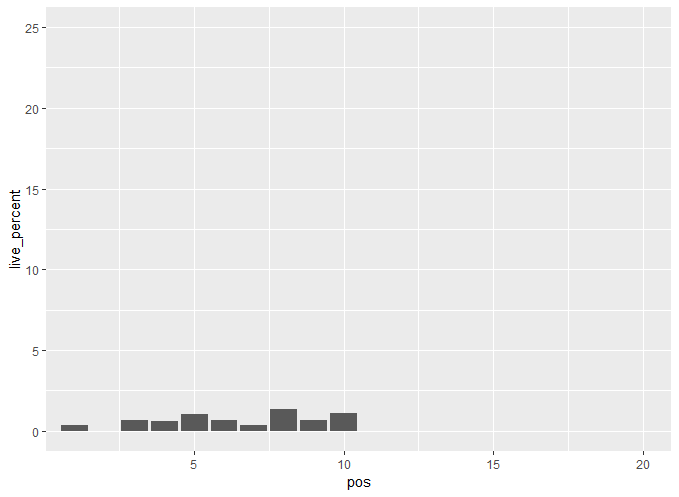
Length 8 – 16 live, 1544 die



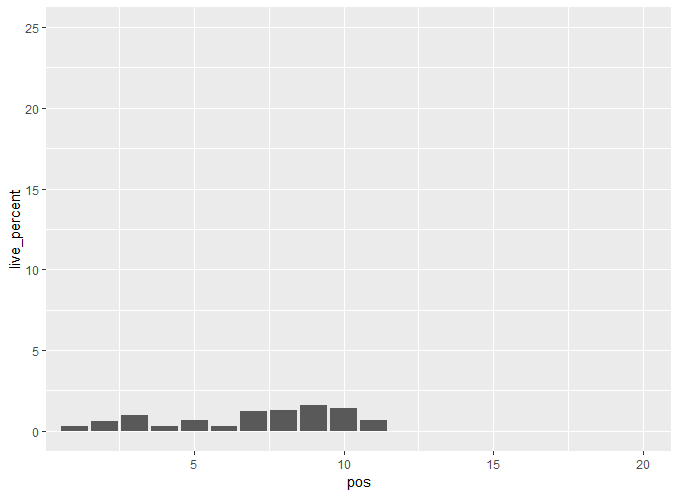
Length 9 – 21 live, 1590 die



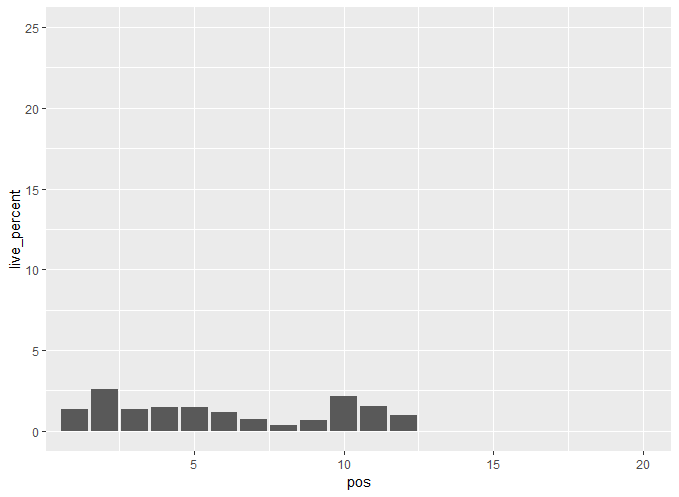
Length 10 – 11 live, 1513 die



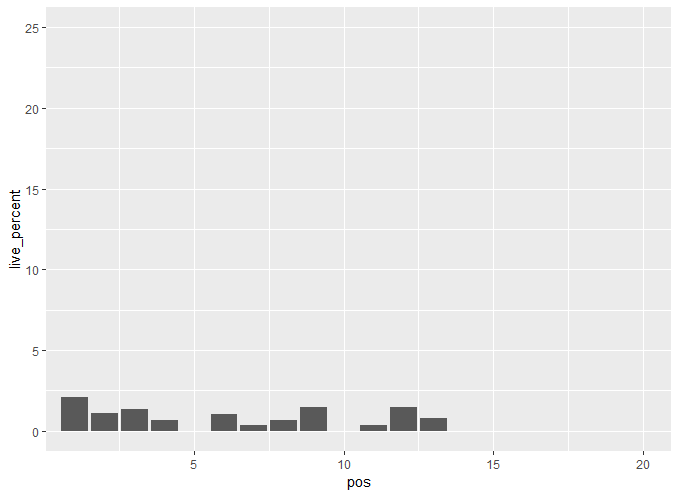
Length 11 – 11 live, 1581 die



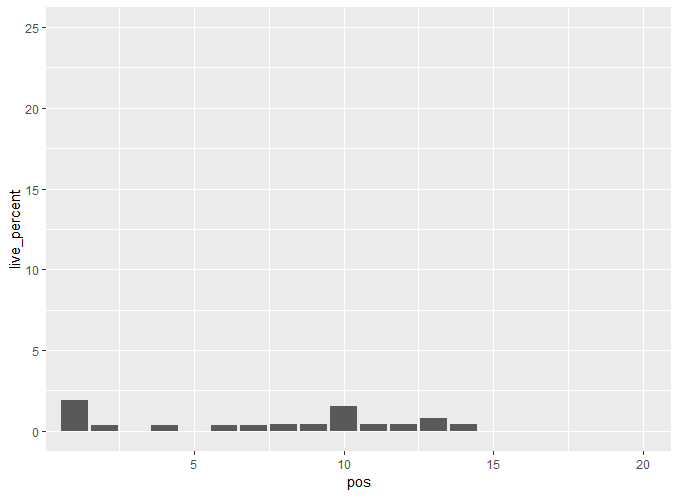
Length 12 – 23 live, 1471 die



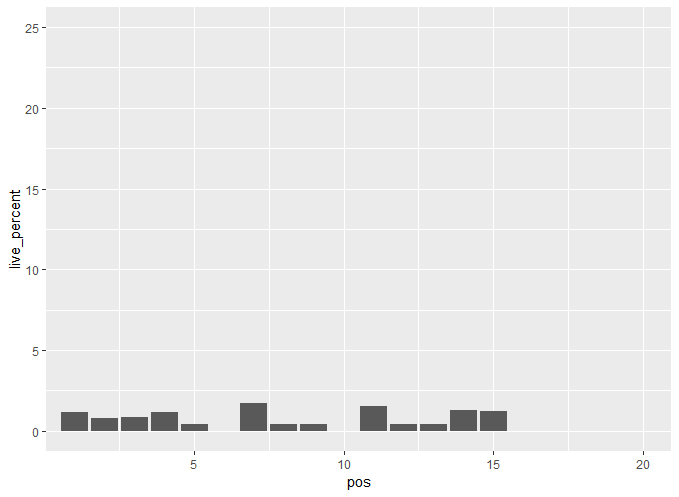
Length 13 – 18 live, 1508 die



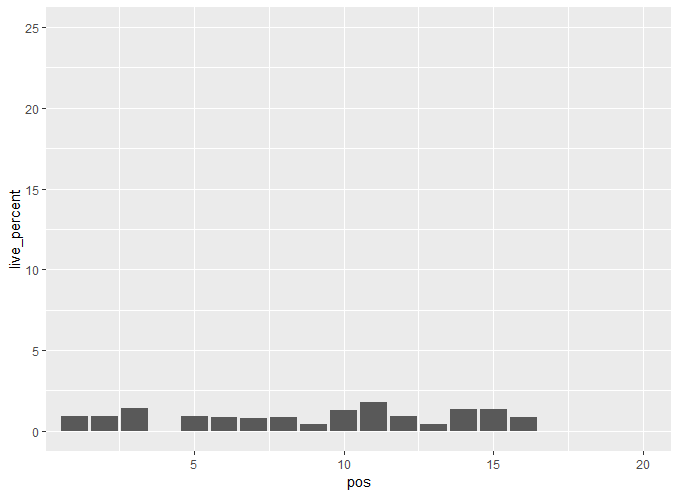
Length 14 – 10 live, 1424 die



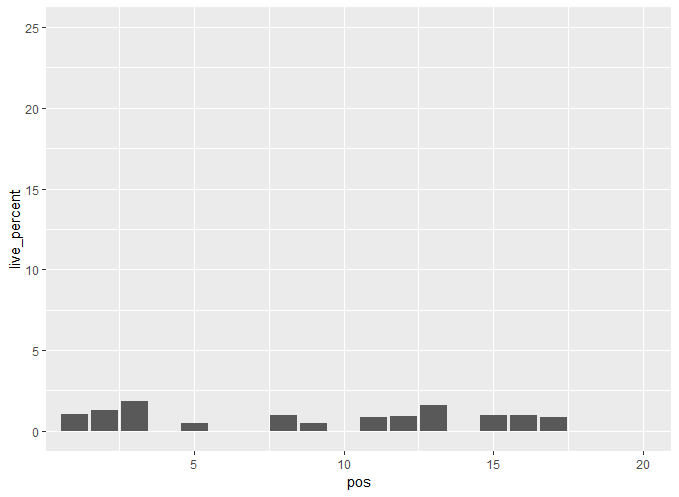
Length 15 – 11 live, 1366 die



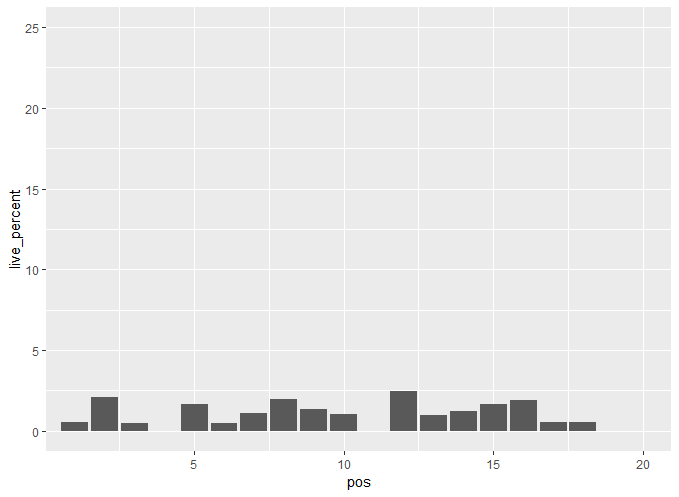
Length 16 – 16 live, 1318 die



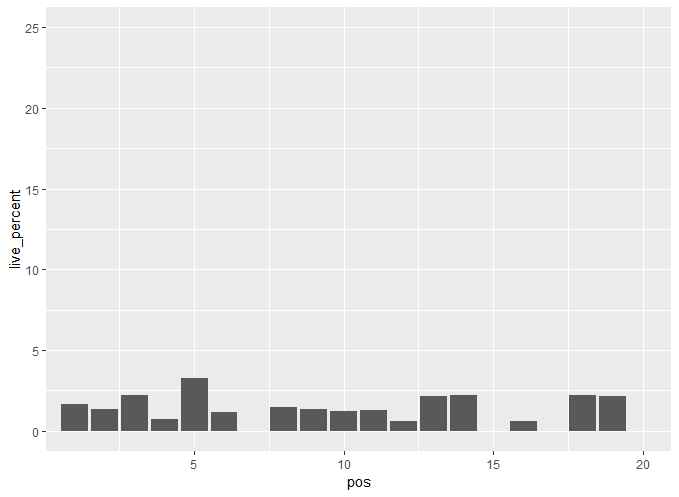
Length 17 – 11 live, 1585 die



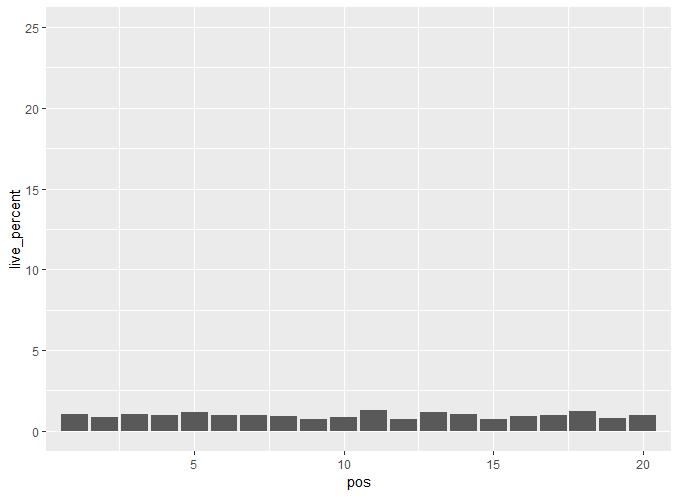
Length 18 – 18 live, 1084 die



Length 19 – 15 live, 896 die



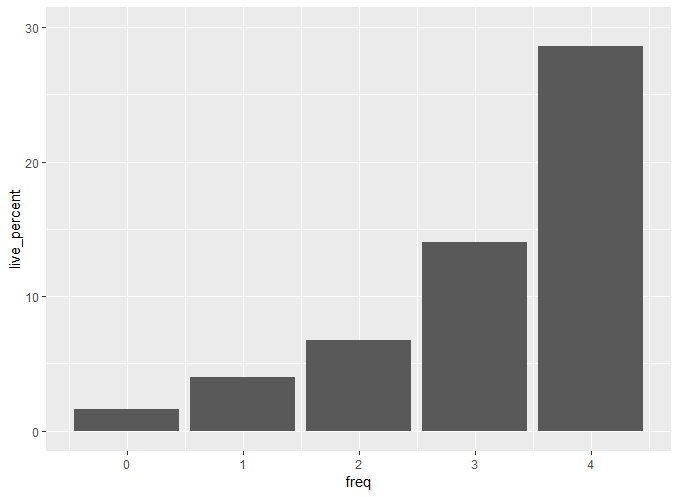
Length 20 – 311 live, 26016 die



Conclusion: not positioning but quantity of basics. Same as aromatic and acidic that is deciding functionality.

1. Increasing frequency of motif

(?=[WYF].{0,4}[DE]|[DE].{0,4}[WYF])



[RHK]

