

Learning a Codebase

Derek Somerville

November 26, 2025

Contents

1 Repository	3
1.1 Glossary - Summary	3
2 Repository: 14	4
2.1 Box plot developer - 14 For packages touched for each period	4
2.2 Percentage touched reached - 14 For component touched for to reach 70.0 percentage known .	10
2.3 Percentage touched reached - 14 For component touched for to reach 80.0 percentage known .	13
2.4 Percentage touched reached - 14 For component touched for to reach 90.0 percentage known .	16
3 Repository: 21	19
3.1 Box plot developer - 21 For packages touched for each period	19
3.2 Percentage touched reached - 21 For component touched for to reach 70.0 percentage known .	25
3.3 Percentage touched reached - 21 For component touched for to reach 80.0 percentage known .	28
3.4 Percentage touched reached - 21 For component touched for to reach 90.0 percentage known .	31
3.5 Smooth moving average - 14 For packages touched for each unit	34
3.6 Smooth moving average - 21 For packages touched for each unit	46
4 Repository pull request - Pull requests	58

1 Repository

1.1 Glossary - Summary

- The founder developer starts in the first six months of a project.
- Late joiner developers begin after six months.
- Transient developers have ten (10) or fewer commits.
- Moderate developers have more than ten (10) commits but fewer than 50 commits or commits for less than 250 days.
- Sustained developers make 50 or more commits and commit for 250 days or more.

2 Repository: 14

2.1 Box plot developer - 14 For packages touched for each period

A box plot of packages touched on average each period the number of commits to components touched.

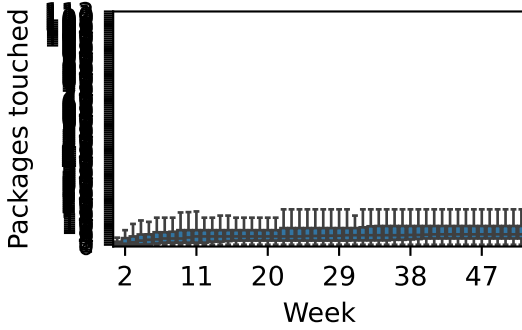
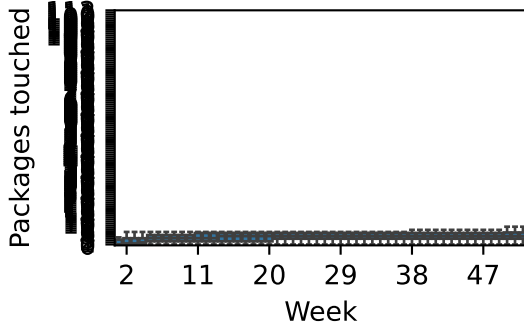
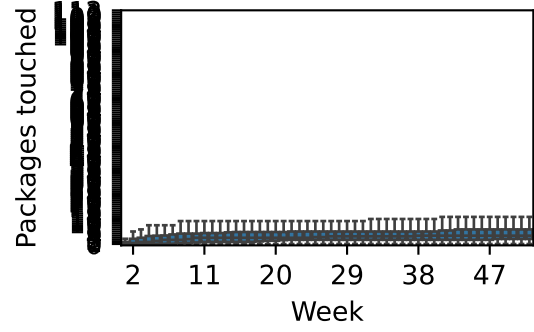


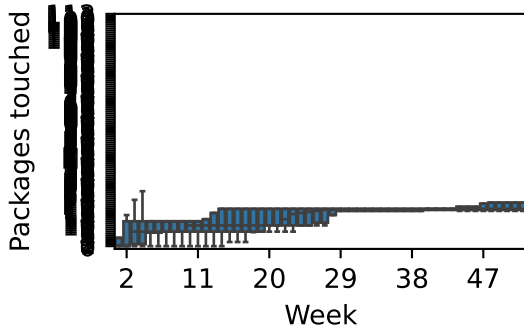
Figure 1: A time series of the number of weeks (x-axis) against the average (mean) total packages touched (y-axis) for all developers 154.



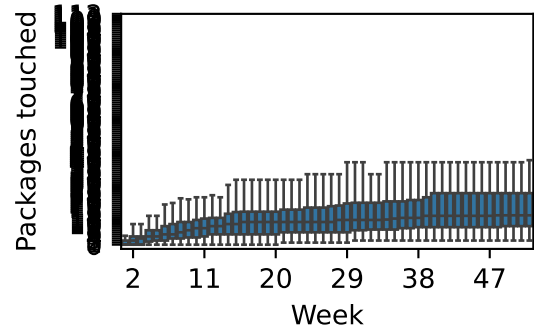
(a) Transient founder developers (8) showing packages touched



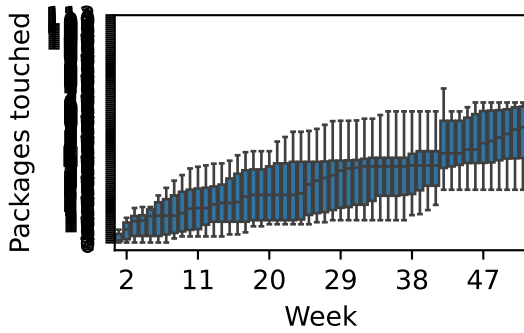
(b) Transient later joiner developers (119) showing packages touched



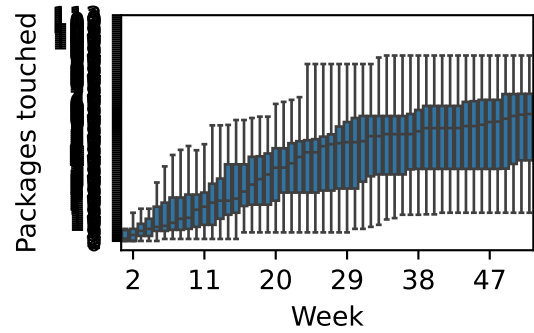
(c) Moderate founder developers (5) showing packages touched



(d) Moderate later joiner developers (85) showing packages touched



(e) Sustained founder developers (6) showing packages touched



(f) Sustained later joiner developers (21) showing packages touched

Figure 2: A box plot of total packages touched on mean each month, with quartile shading.

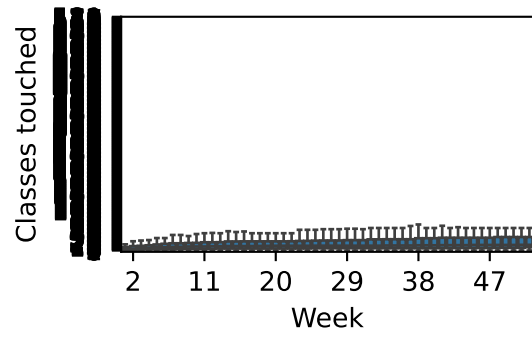
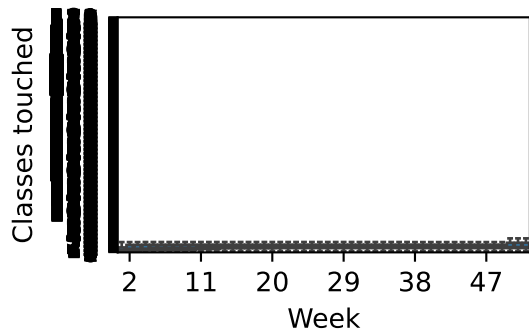
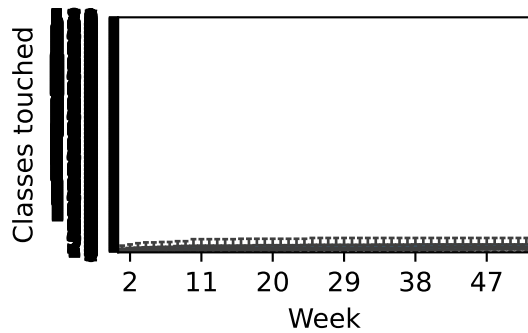


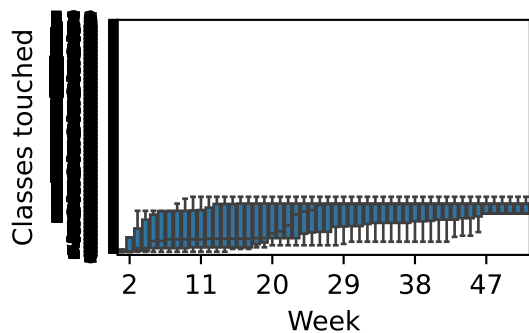
Figure 3: A time series of the number of weeks (x-axis) against the average (mean) total classes touched (y-axis) for all developers 154.



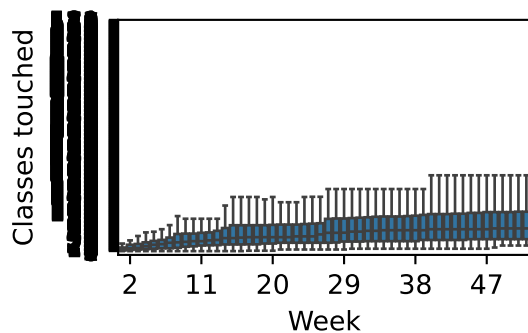
(a) Transient founder developers (8) showing classes touched



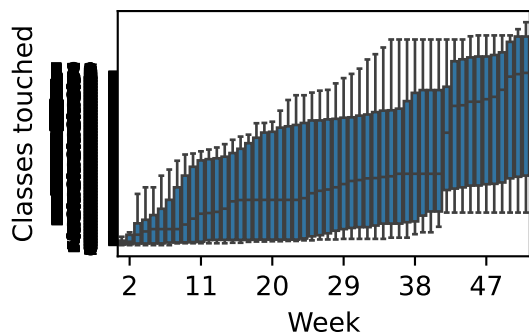
(b) Transient later joiner developers (119) showing classes touched



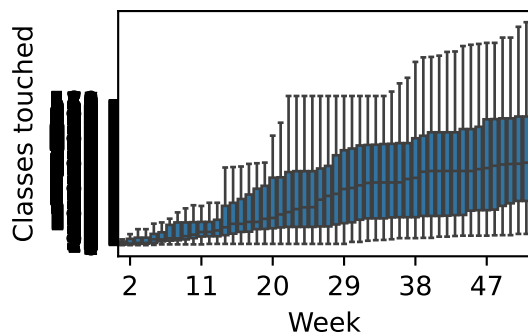
(c) Moderate founder developers (5) showing classes touched



(d) Moderate later joiner developers (85) showing classes touched



(e) Sustained founder developers (6) showing classes touched



(f) Sustained later joiner developers (21) showing classes touched

Figure 4: A box plot of total classes touched on mean each month, with quartile shading.

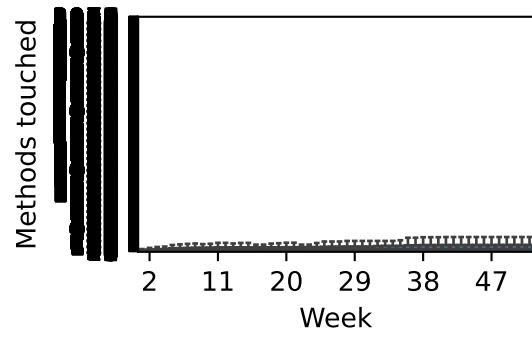
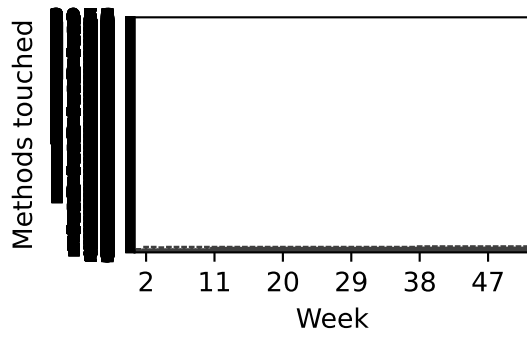
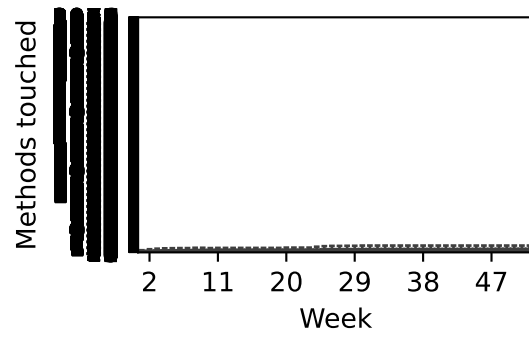


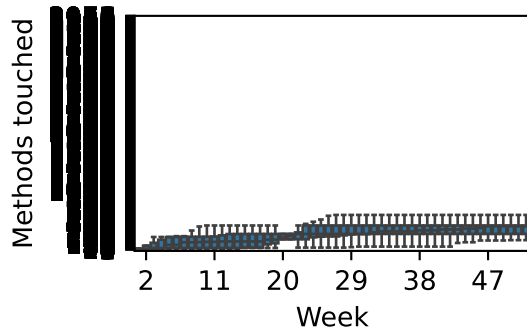
Figure 5: A time series of the number of weeks (x-axis) against the average (mean) total methods touched (y-axis) for all developers 154.



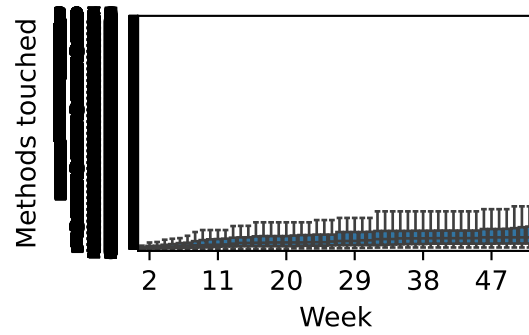
(a) Transient founder developers (8) showing methods touched



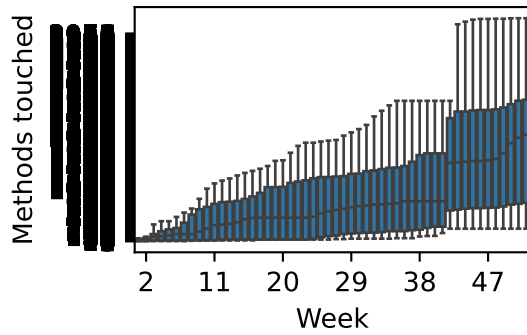
(b) Transient later joiner developers (119) showing methods touched



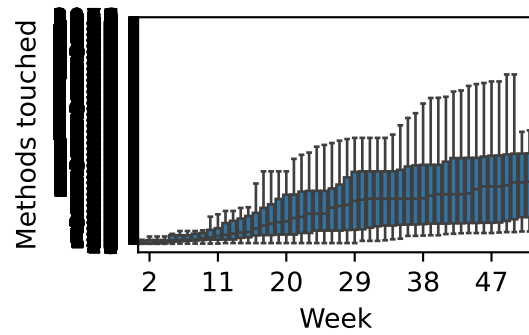
(c) Moderate founder developers (5) showing methods touched



(d) Moderate later joiner developers (85) showing methods touched



(e) Sustained founder developers (6) showing methods touched

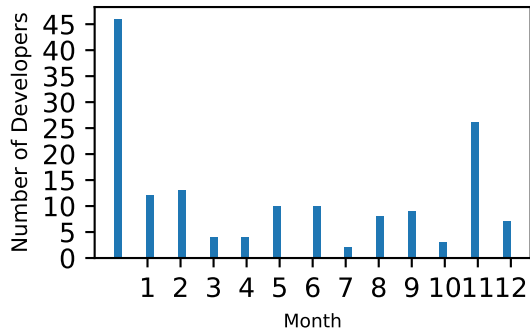


(f) Sustained later joiner developers (21) showing methods touched

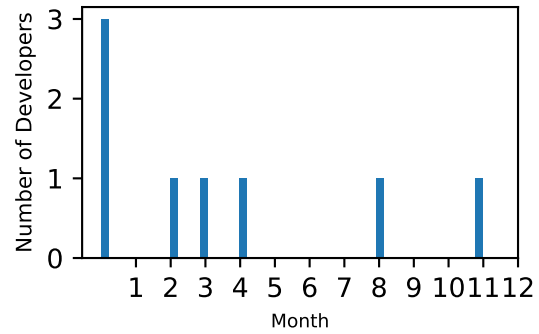
Figure 6: A box plot of total methods touched on mean each month, with quartile shading.

2.2 Percentage touched reached - 14 For component touched for to reach 70.0 percentage known

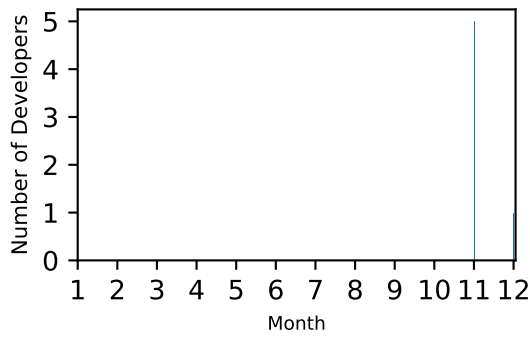
A histogram of number of developers that have touched a percentage of packages each period



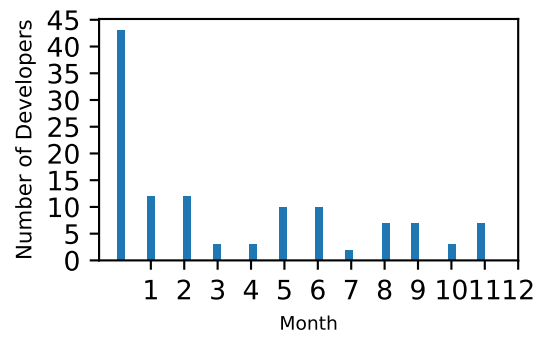
(a) All



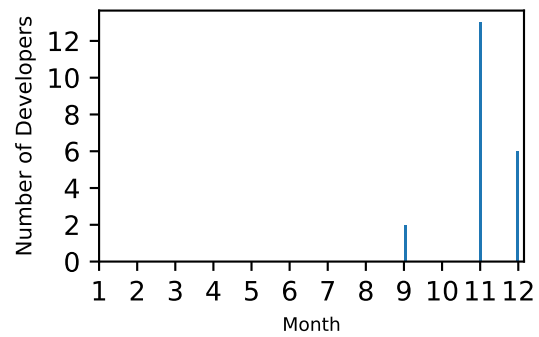
(b) transient founder



(c) sustained founder

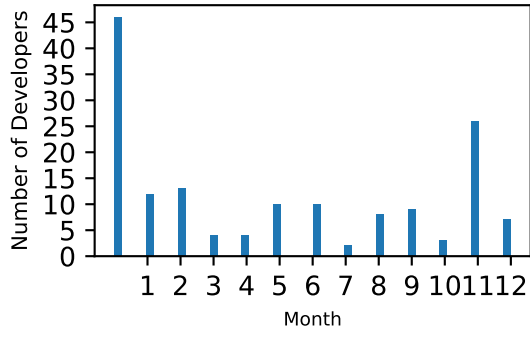


(d) transient later joiner

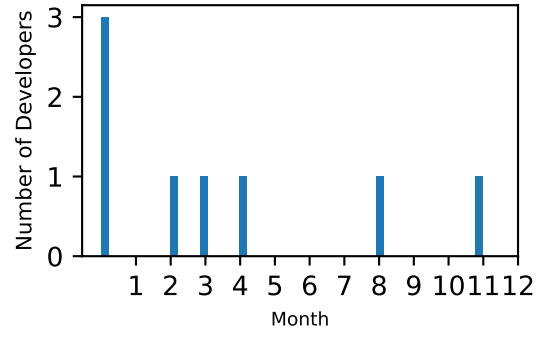


(e) sustained later joiner

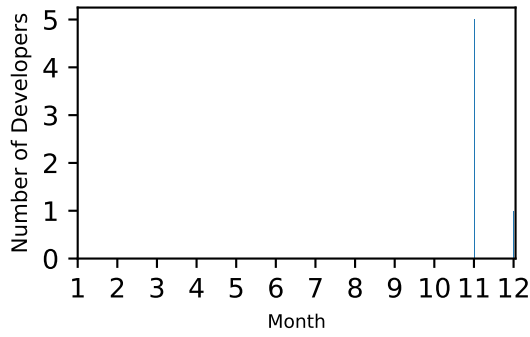
Figure 7: A histogram of number of developers that have touched a percentage (70.0%) of packages each month of a year.



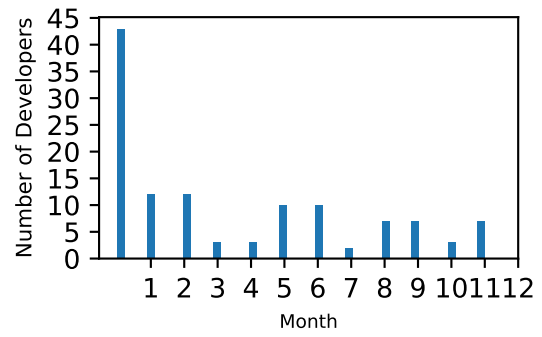
(a) All



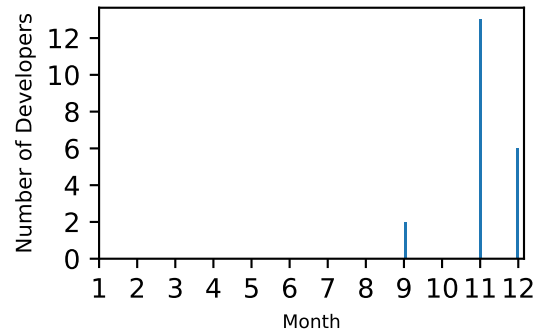
(b) transient founder



(c) sustained founder

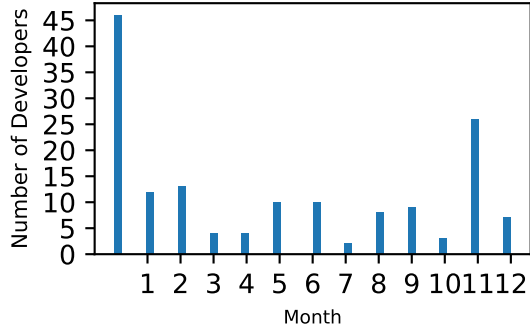


(d) transient later joiner

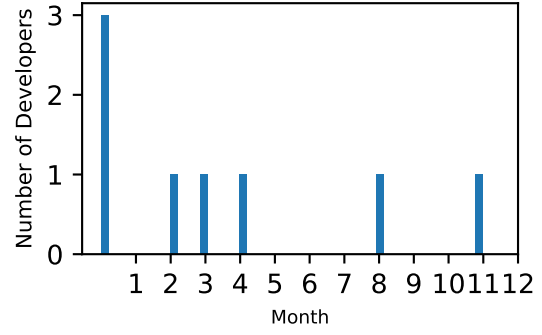


(e) sustained later joiner

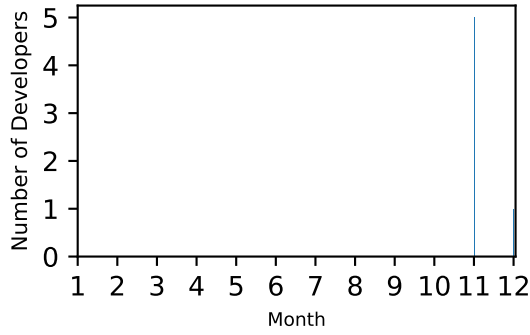
Figure 8: A histogram of number of developers that have touched a percentage (70.0%) of classes each month of a year.



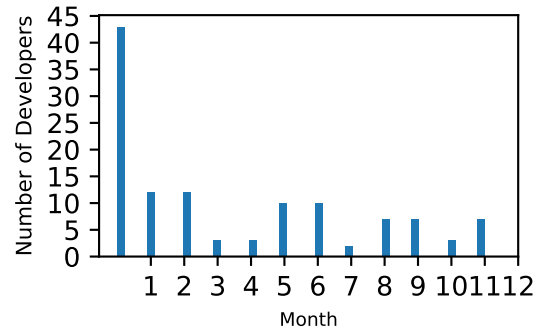
(a) All



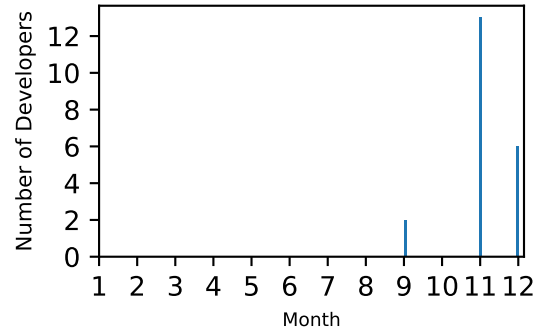
(b) transient founder



(c) sustained founder



(d) transient later joiner

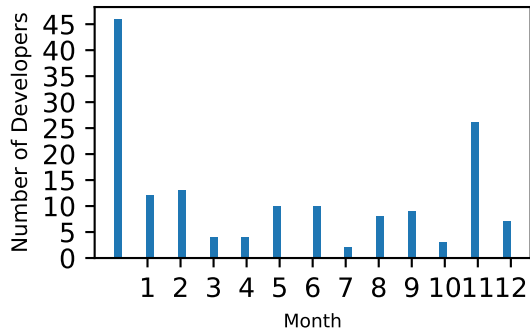


(e) sustained later joiner

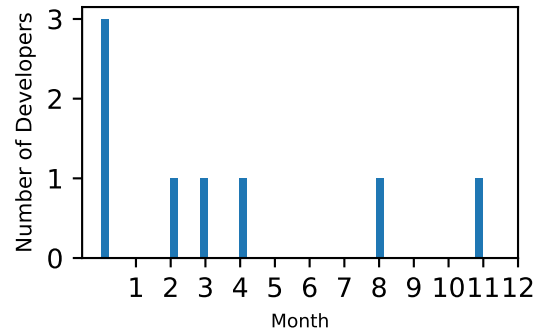
Figure 9: A histogram of number of developers that have touched a percentage (70.0%) of methods each month of a year.

2.3 Percentage touched reached - 14 For component touched for to reach 80.0 percentage known

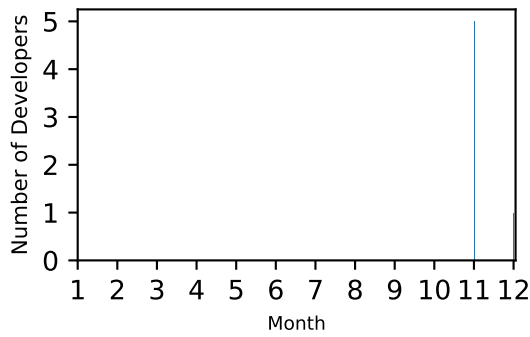
A histogram of number of developers that have touched a percentage of packages each period



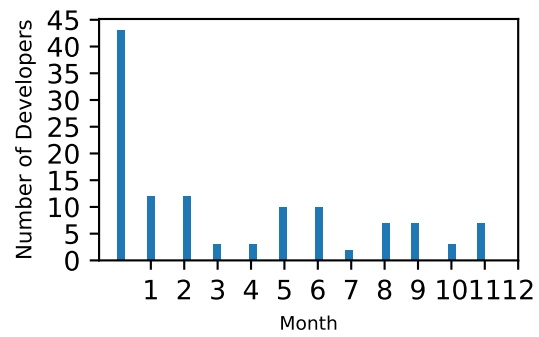
(a) All



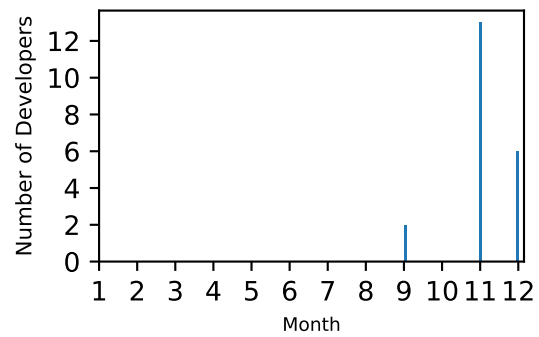
(b) transient founder



(c) sustained founder

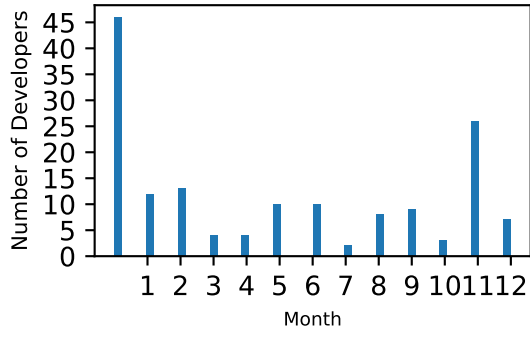


(d) transient later joiner

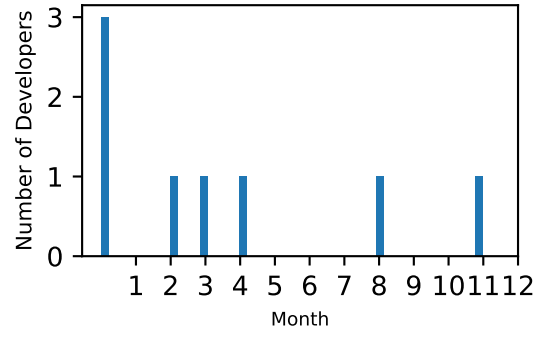


(e) sustained later joiner

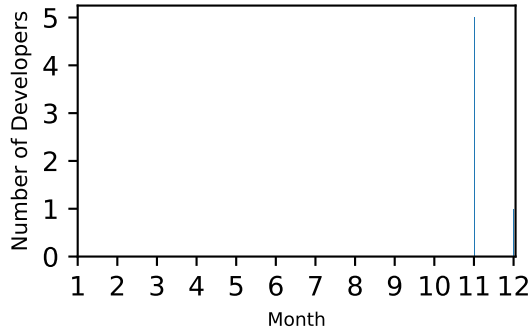
Figure 10: A histogram of number of developers that have touched a percentage (80.0%) of packages each month of a year.



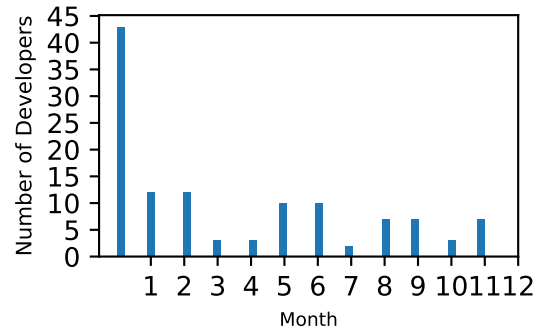
(a) All



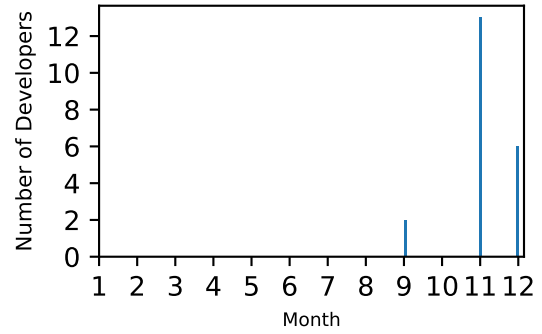
(b) transient founder



(c) sustained founder

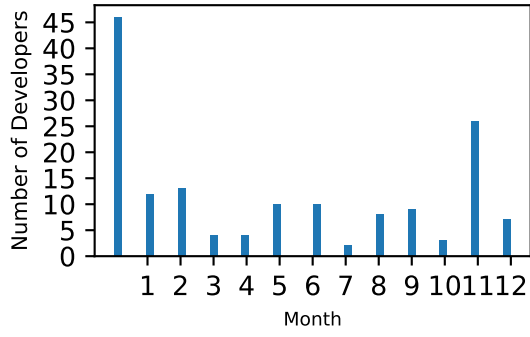


(d) transient later joiner

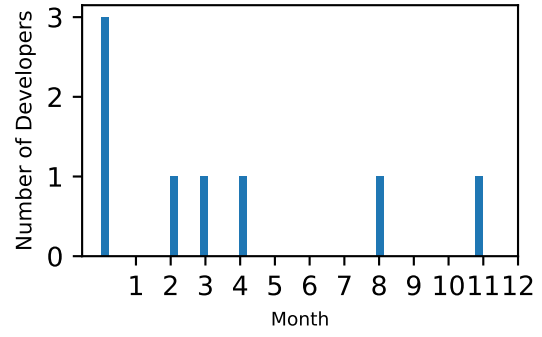


(e) sustained later joiner

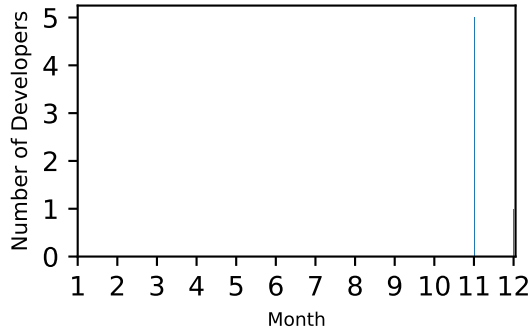
Figure 11: A histogram of number of developers that have touched a percentage (80.0%) of classes each month of a year.



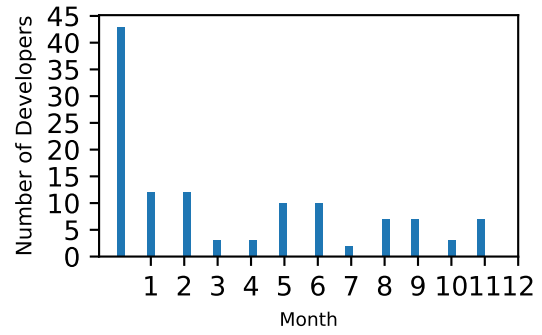
(a) All



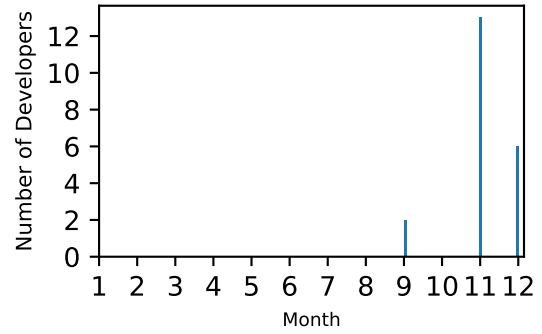
(b) transient founder



(c) sustained founder



(d) transient later joiner

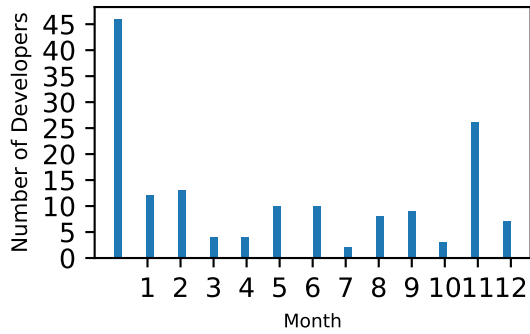


(e) sustained later joiner

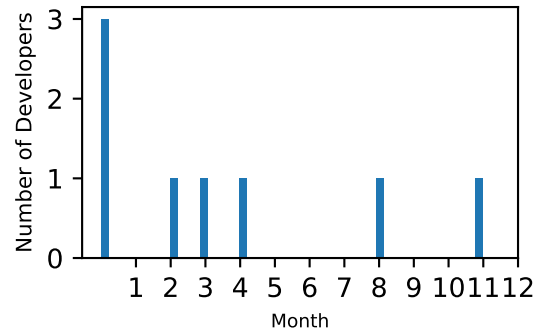
Figure 12: A histogram of number of developers that have touched a percentage (80.0%) of methods each month of a year.

2.4 Percentage touched reached - 14 For component touched for to reach 90.0 percentage known

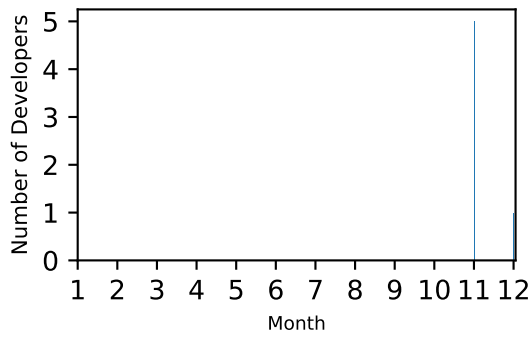
A histogram of number of developers that have touched a percentage of packages each period



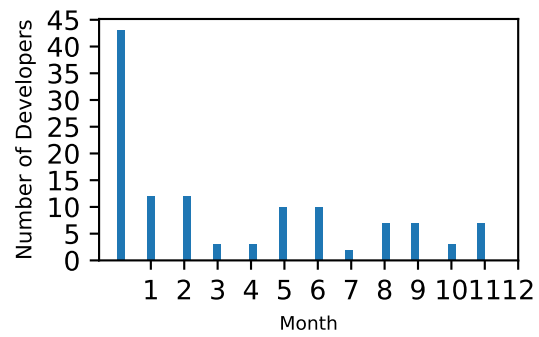
(a) All



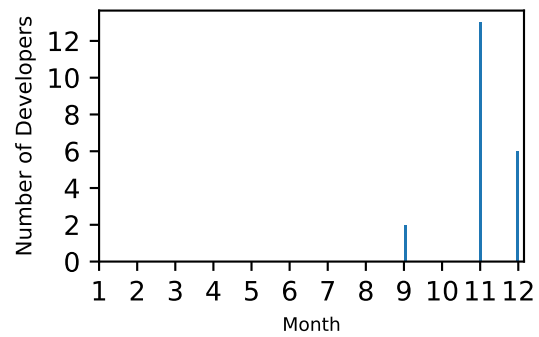
(b) transient founder



(c) sustained founder

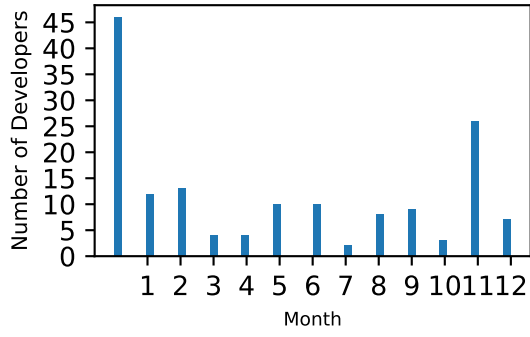


(d) transient later joiner

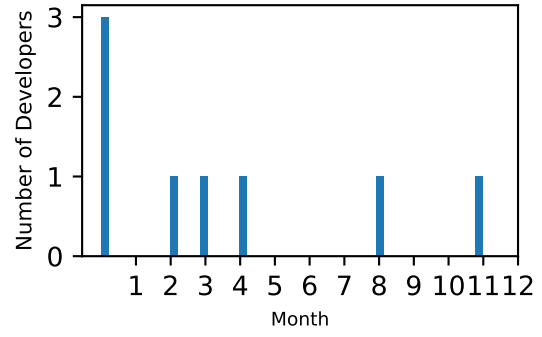


(e) sustained later joiner

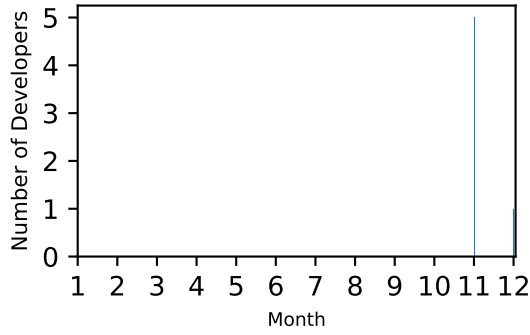
Figure 13: A histogram of number of developers that have touched a percentage (90.0%) of packages each month of a year.



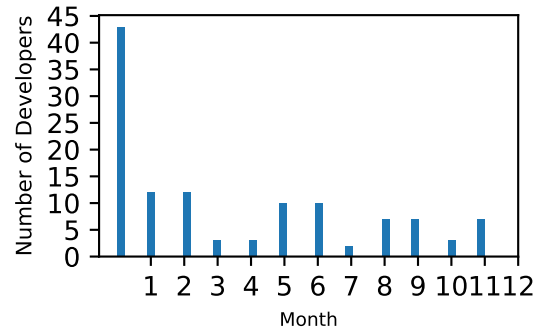
(a) All



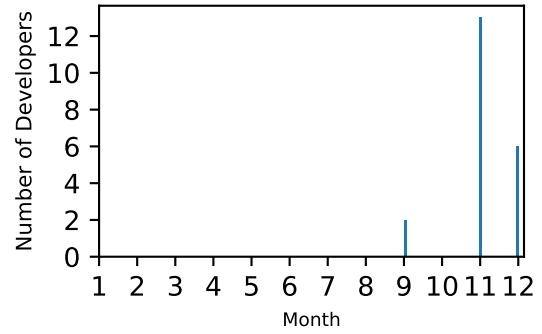
(b) transient founder



(c) sustained founder

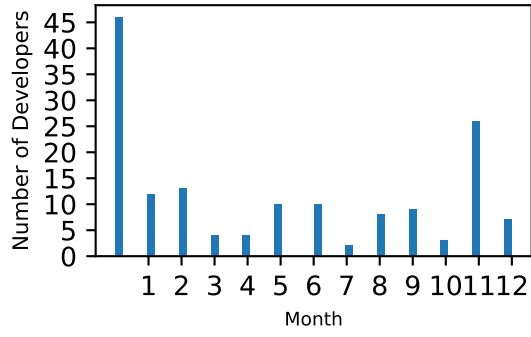


(d) transient later joiner

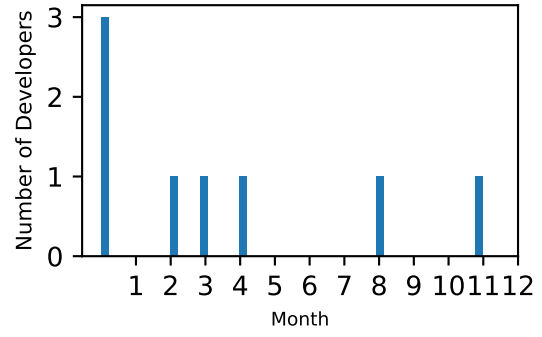


(e) sustained later joiner

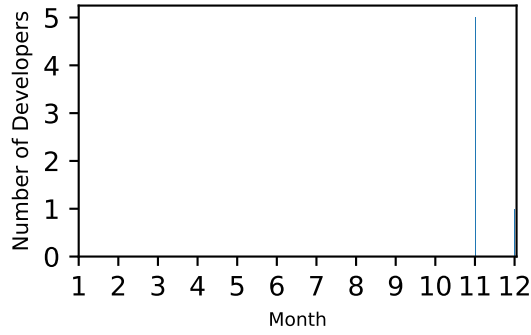
Figure 14: A histogram of number of developers that have touched a percentage (90.0%) of classes each month of a year.



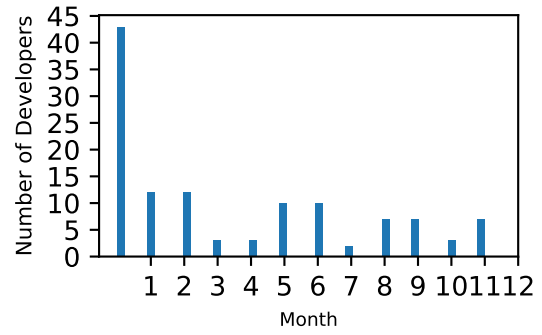
(a) All



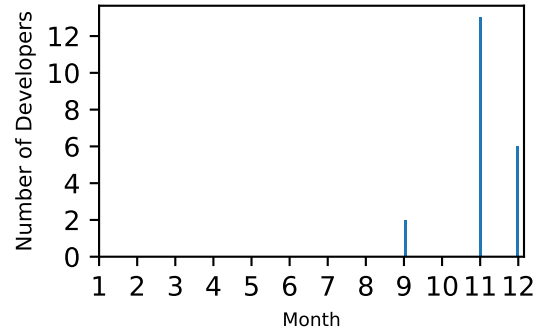
(b) transient founder



(c) sustained founder



(d) transient later joiner



(e) sustained later joiner

Figure 15: A histogram of number of developers that have touched a percentage (90.0%) of methods each month of a year.

3 Repository: 21

3.1 Box plot developer - 21 For packages touched for each period

A box plot of packages touched on average each period the number of commits to components touched.

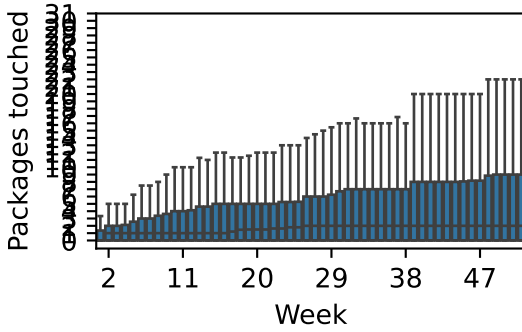
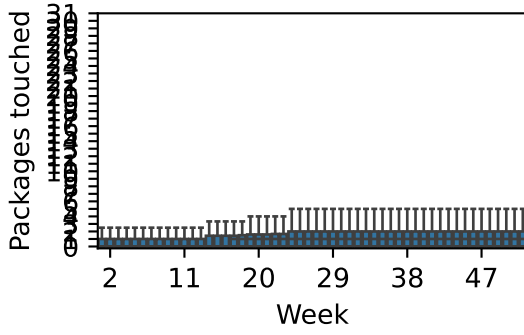
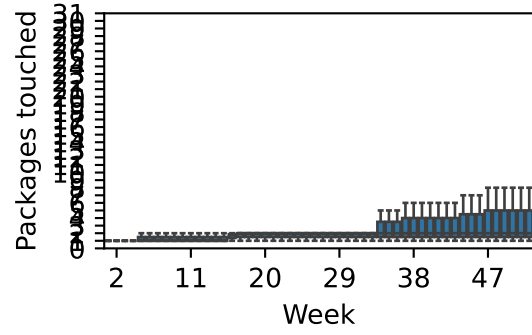


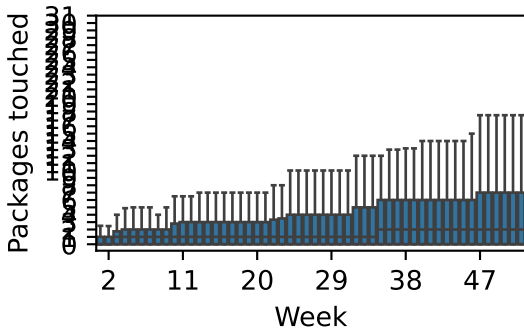
Figure 16: A time series of the number of weeks (x-axis) against the average (mean) total packages touched (y-axis) for all developers 204.



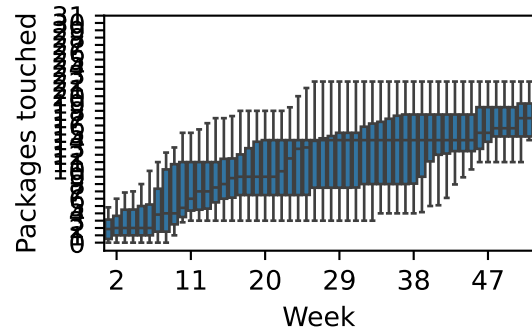
(a) Transient later joiner developers (113) showing packages touched



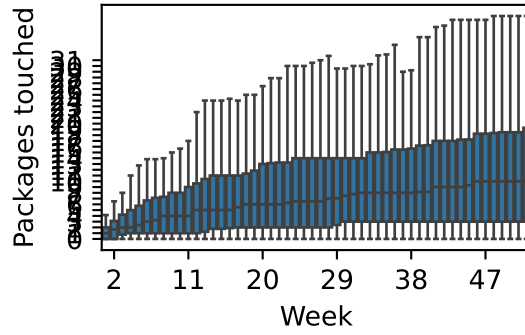
(b) Moderate founder developers (3) showing packages touched



(c) Moderate later joiner developers (93) showing packages touched



(d) Sustained founder developers (7) showing packages touched



(e) Sustained later joiner developers (84) showing packages touched

Figure 17: A box plot of total packages touched on mean each month, with quartile shading.

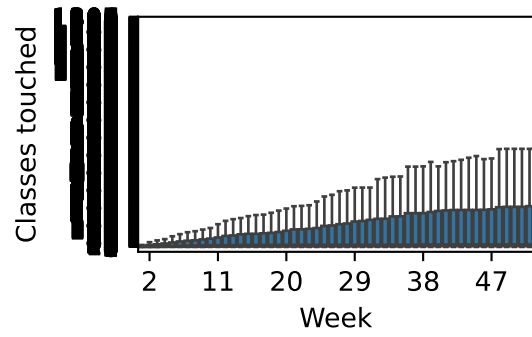
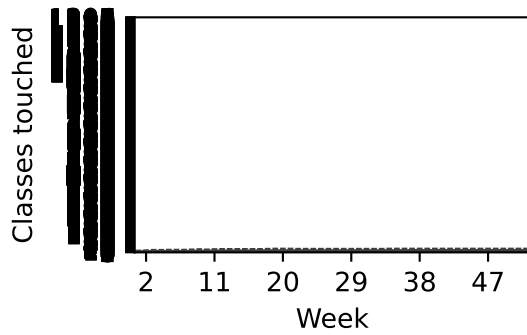
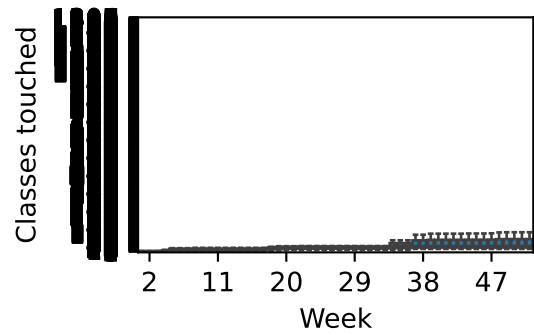


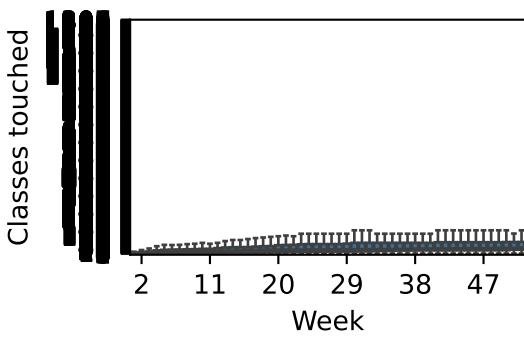
Figure 18: A time series of the number of weeks (x-axis) against the average (mean) total classes touched (y-axis) for all developers 204.



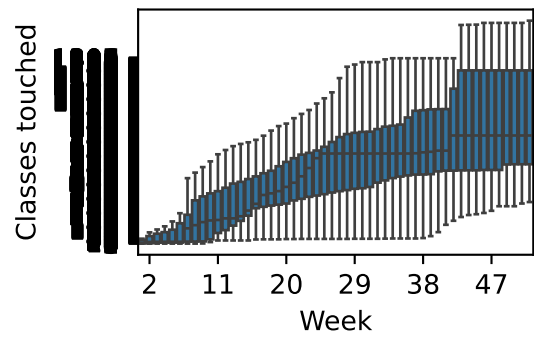
(a) Transient later joiner developers (113) showing classes touched



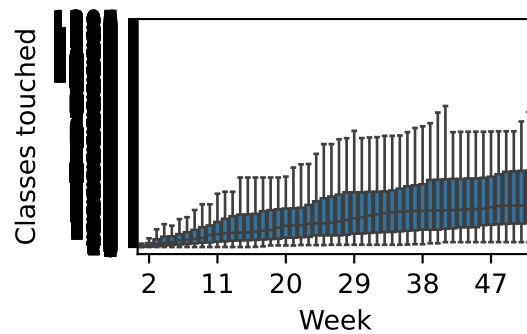
(b) Moderate founder developers (3) showing classes touched



(c) Moderate later joiner developers (93) showing classes touched



(d) Sustained founder developers (7) showing classes touched



(e) Sustained later joiner developers (84) showing classes touched

Figure 19: A box plot of total classes touched on mean each month, with quartile shading.

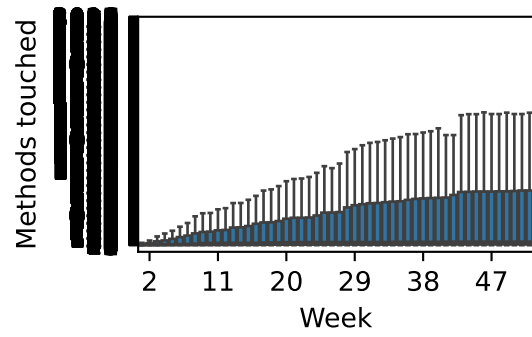
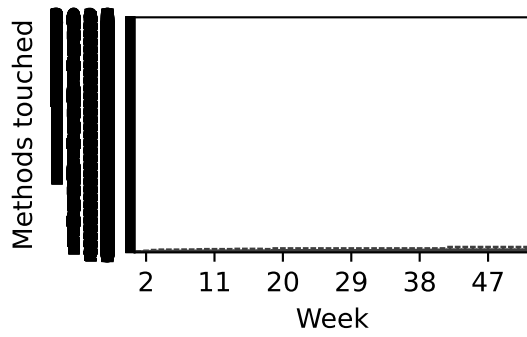
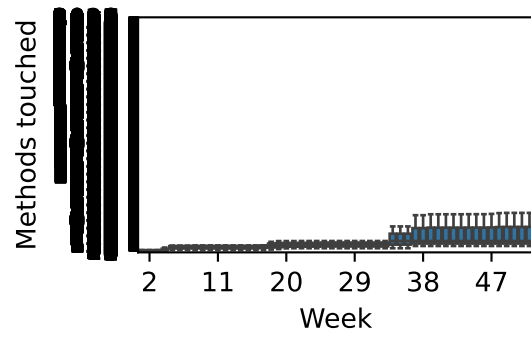


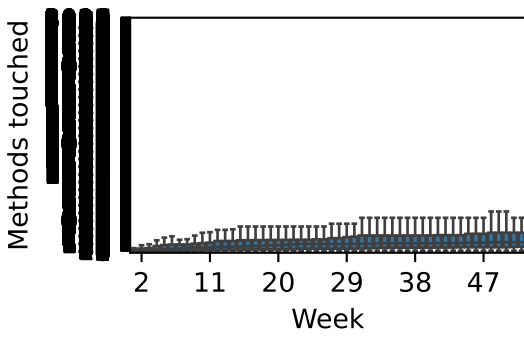
Figure 20: A time series of the number of weeks (x-axis) against the average (mean) total methods touched (y-axis) for all developers 204.



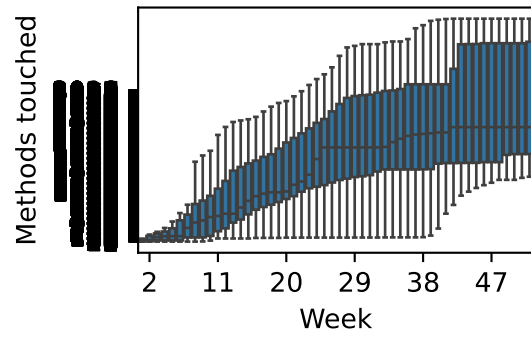
(a) Transient later joiner developers (113) showing methods touched



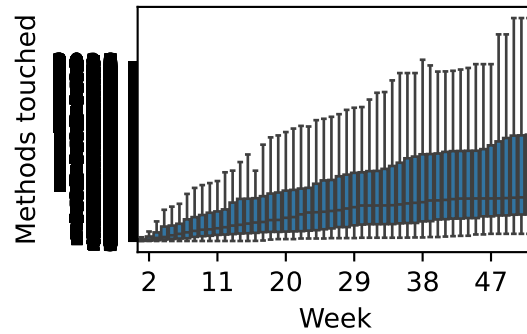
(b) Moderate founder developers (3) showing methods touched



(c) Moderate later joiner developers (93) showing methods touched



(d) Sustained founder developers (7) showing methods touched

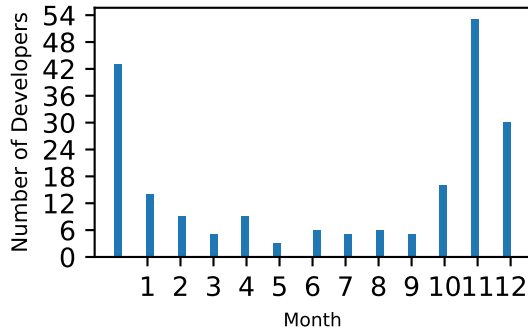


(e) Sustained later joiner developers (84) showing methods touched

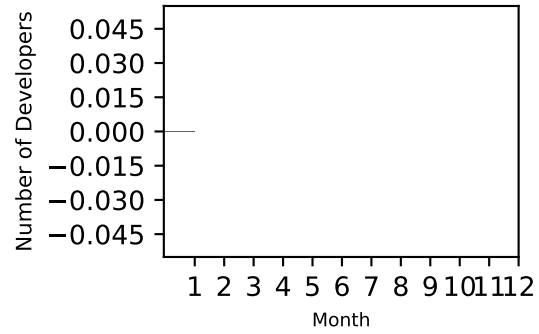
Figure 21: A box plot of total methods touched on mean each month, with quartile shading.

3.2 Percentage touched reached - 21 For component touched for to reach 70.0 percentage known

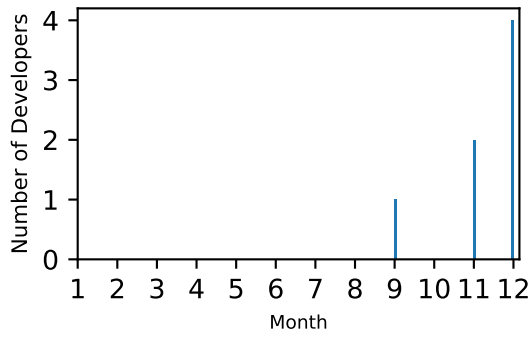
A histogram of number of developers that have touched a percentage of packages each period



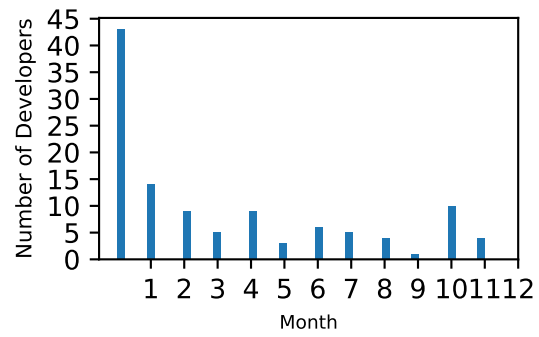
(a) All



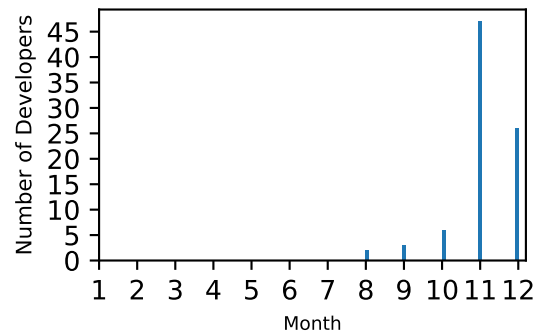
(b) transient founder



(c) sustained founder

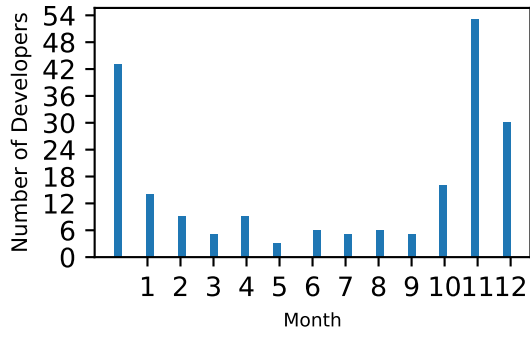


(d) transient later joiner

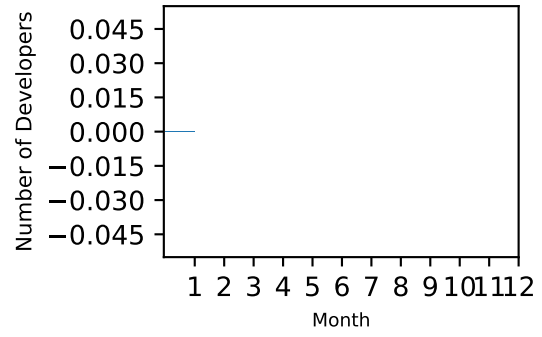


(e) sustained later joiner

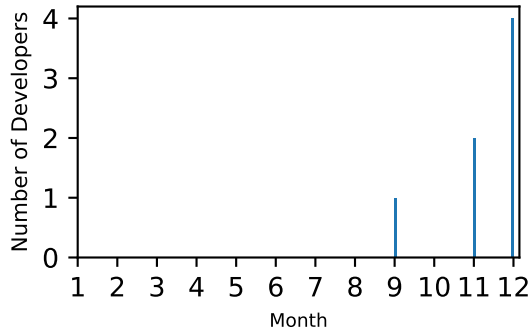
Figure 22: A histogram of number of developers that have touched a percentage (70.0%) of packages each month of a year.



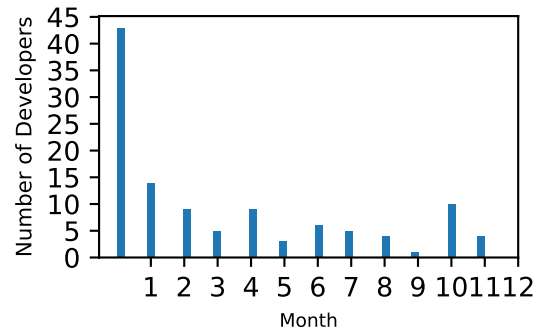
(a) All



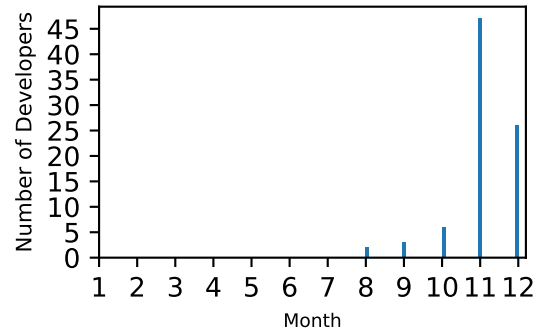
(b) transient founder



(c) sustained founder

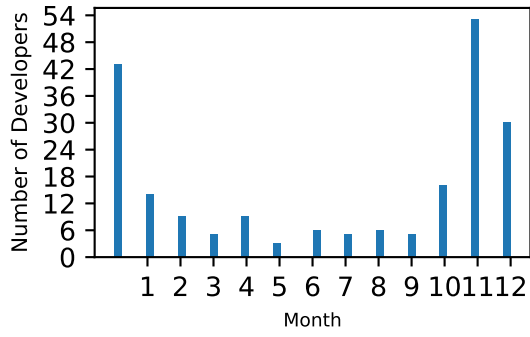


(d) transient later joiner

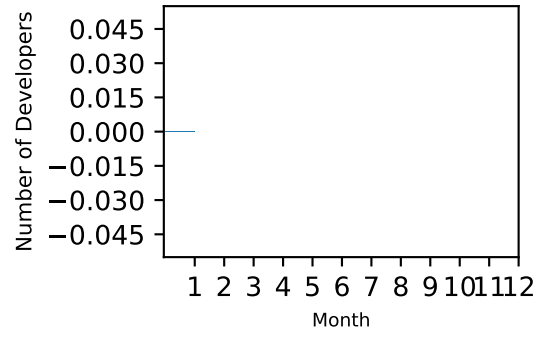


(e) sustained later joiner

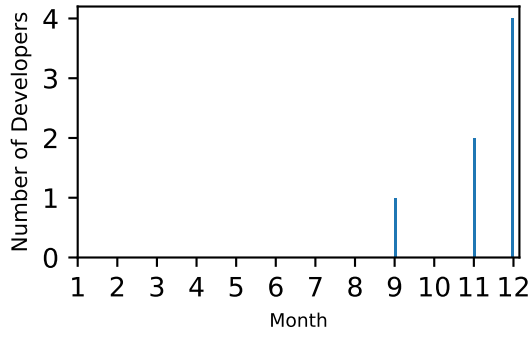
Figure 23: A histogram of number of developers that have touched a percentage (70.0%) of classes each month of a year.



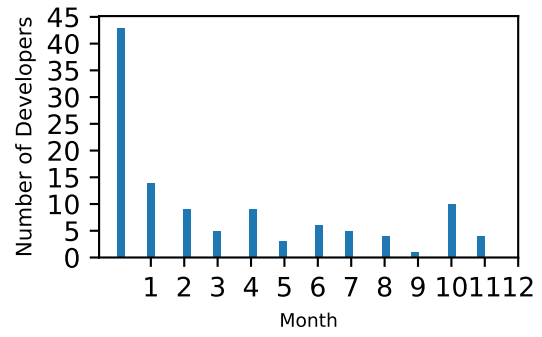
(a) All



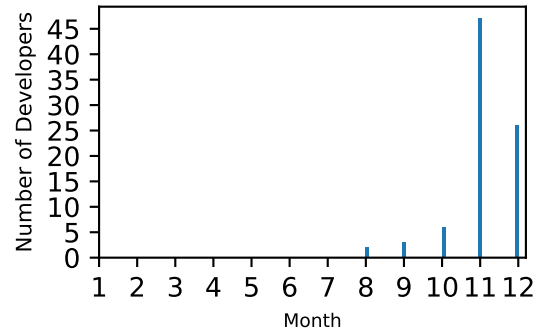
(b) transient founder



(c) sustained founder



(d) transient later joiner

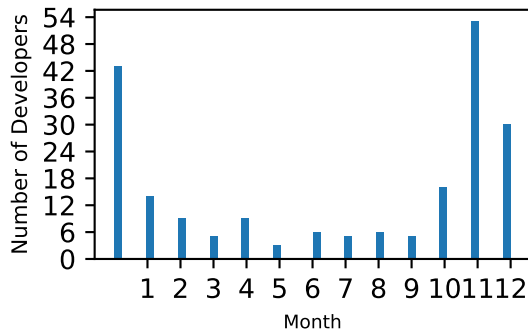


(e) sustained later joiner

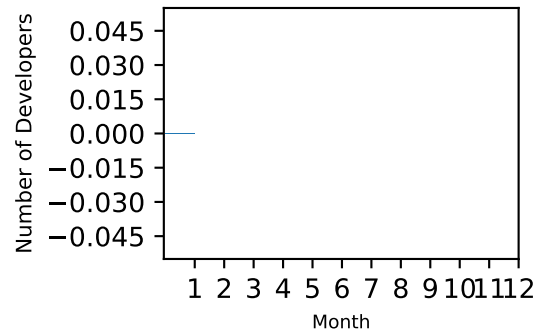
Figure 24: A histogram of number of developers that have touched a percentage (70.0%) of methods each month of a year.

3.3 Percentage touched reached - 21 For component touched for to reach 80.0 percentage known

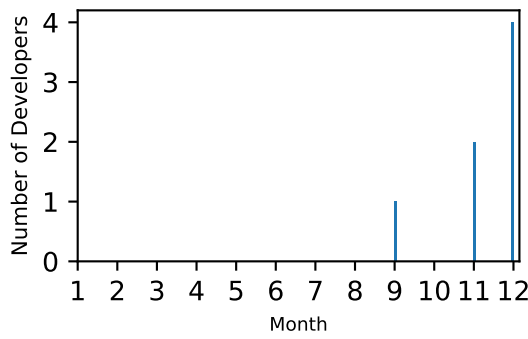
A histogram of number of developers that have touched a percentage of packages each period



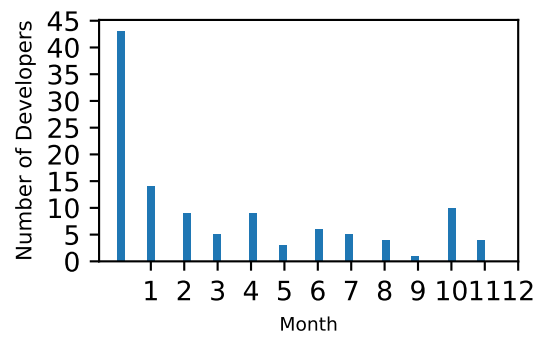
(a) All



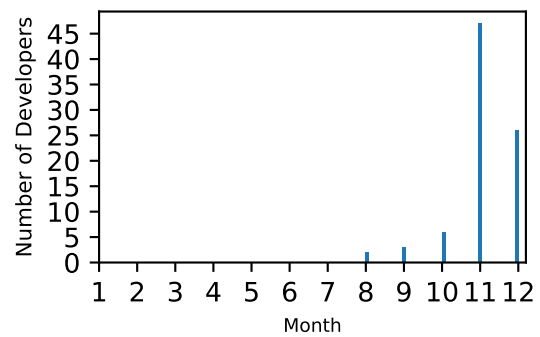
(b) transient founder



(c) sustained founder

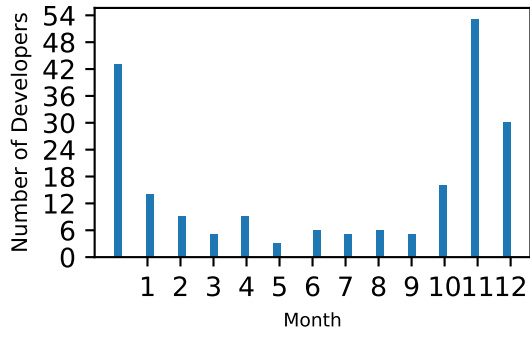


(d) transient later joiner

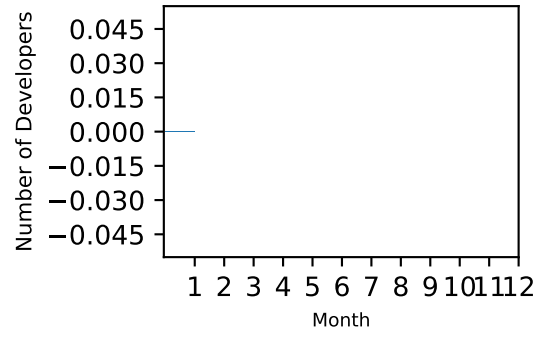


(e) sustained later joiner

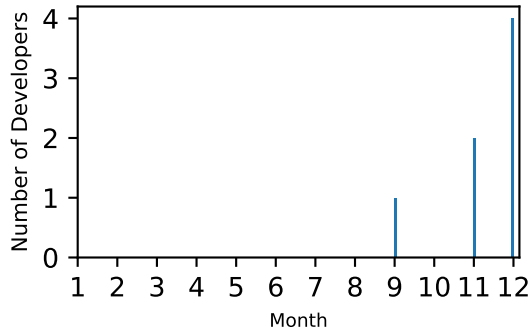
Figure 25: A histogram of number of developers that have touched a percentage (80.0%) of packages each month of a year.



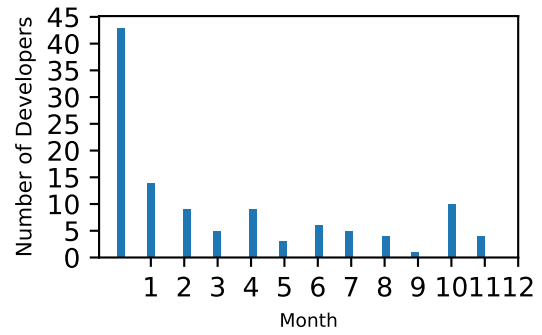
(a) All



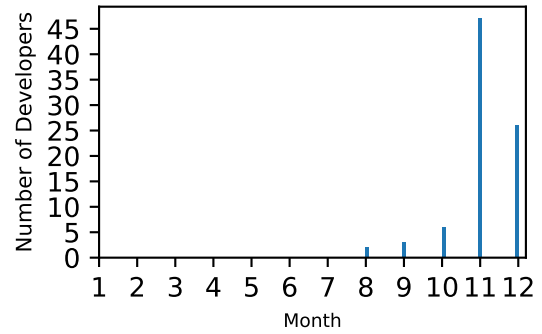
(b) transient founder



(c) sustained founder

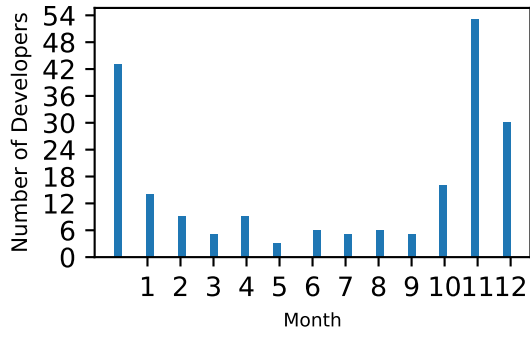


(d) transient later joiner

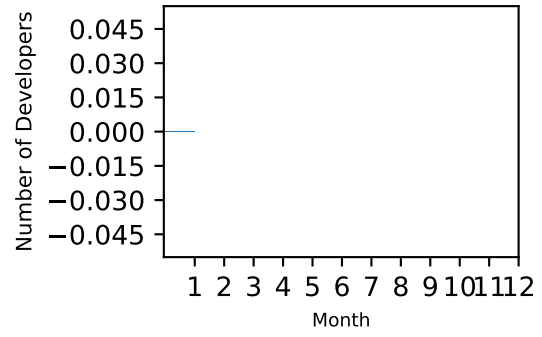


(e) sustained later joiner

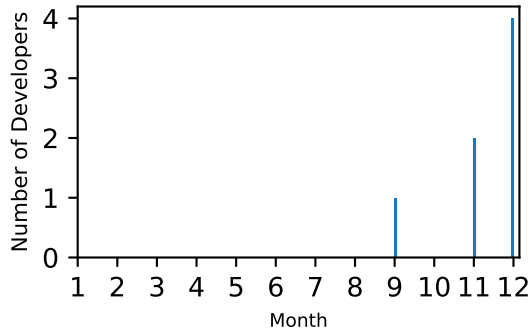
Figure 26: A histogram of number of developers that have touched a percentage (80.0%) of classes each month of a year.



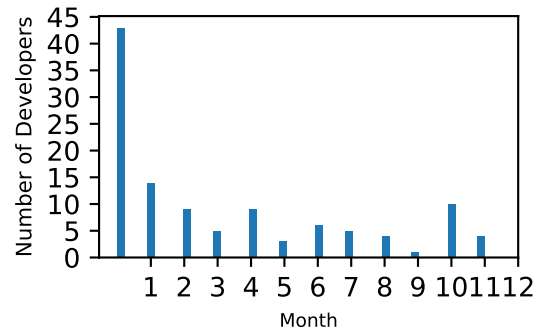
(a) All



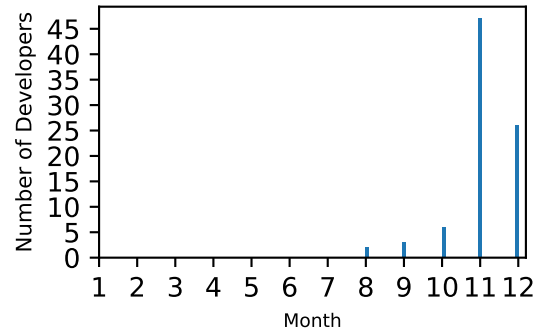
(b) transient founder



(c) sustained founder



(d) transient later joiner

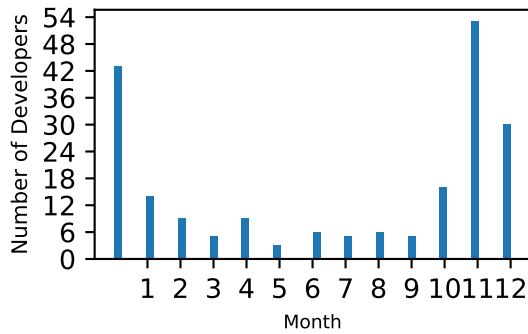


(e) sustained later joiner

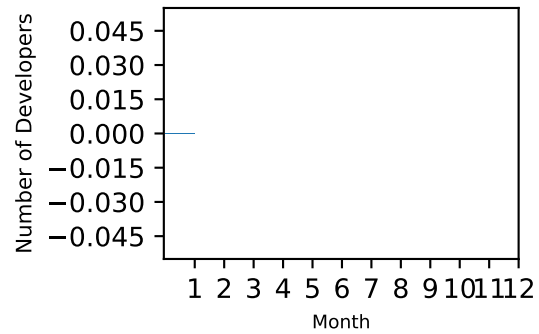
Figure 27: A histogram of number of developers that have touched a percentage (80.0%) of methods each month of a year.

3.4 Percentage touched reached - 21 For component touched for to reach 90.0 percentage known

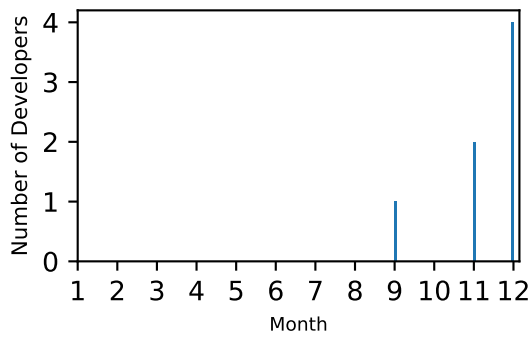
A histogram of number of developers that have touched a percentage of packages each period



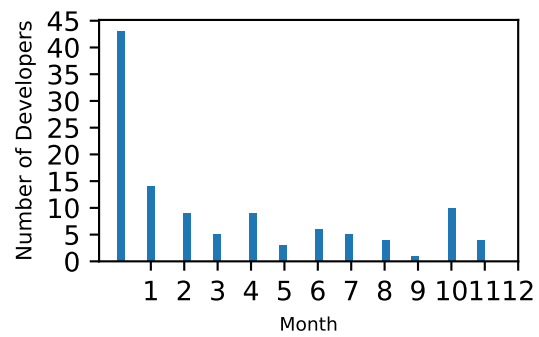
(a) All



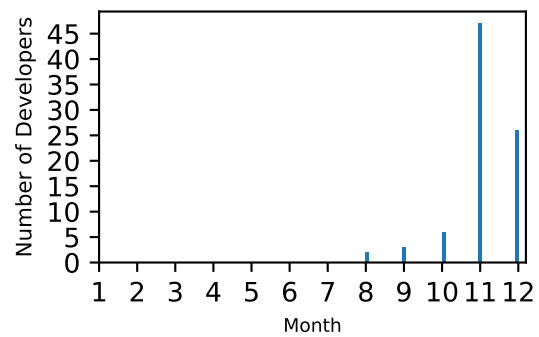
(b) transient founder



(c) sustained founder

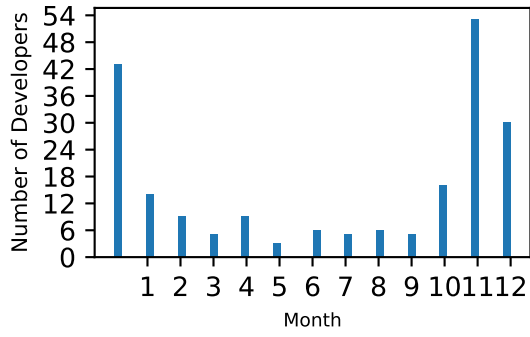


(d) transient later joiner

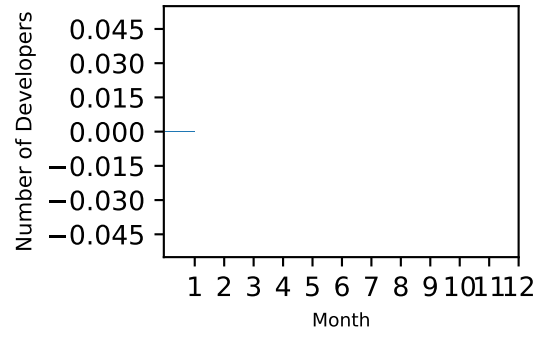


(e) sustained later joiner

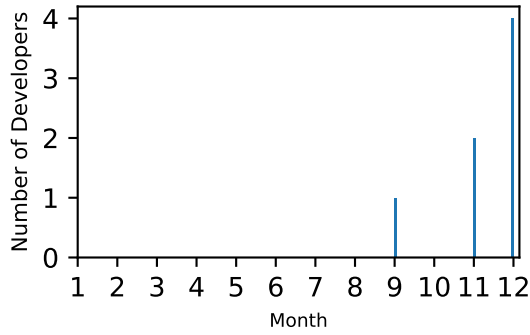
Figure 28: A histogram of number of developers that have touched a percentage (90.0%) of packages each month of a year.



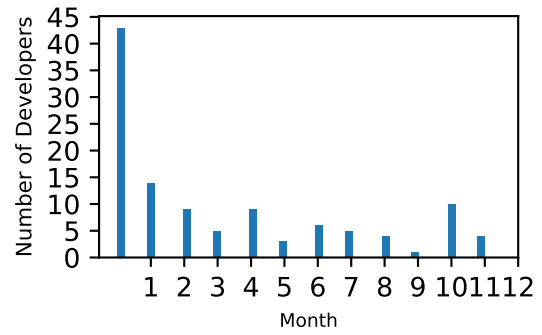
(a) All



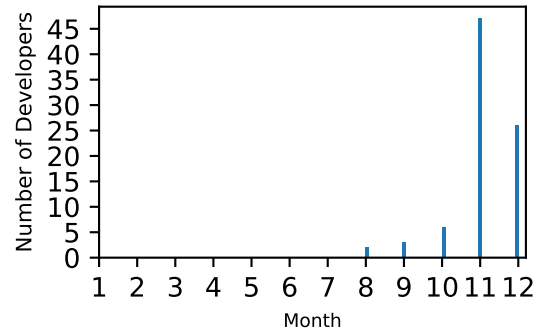
(b) transient founder



(c) sustained founder

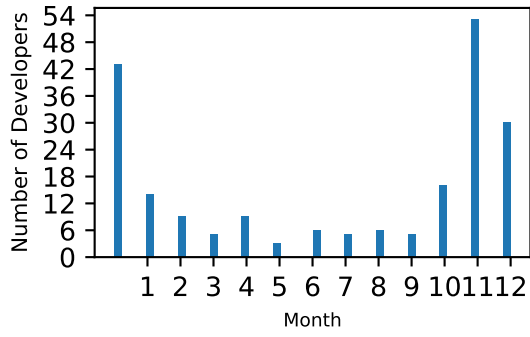


(d) transient later joiner

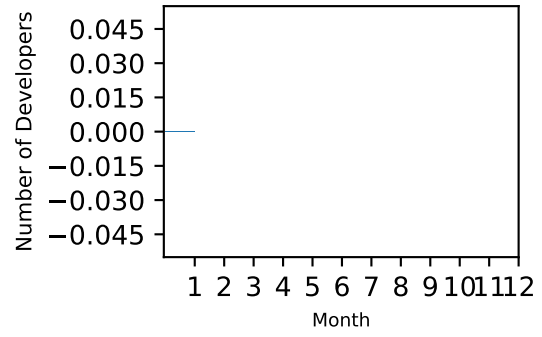


(e) sustained later joiner

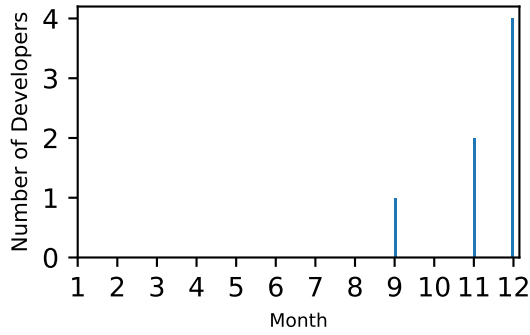
Figure 29: A histogram of number of developers that have touched a percentage (90.0%) of classes each month of a year.



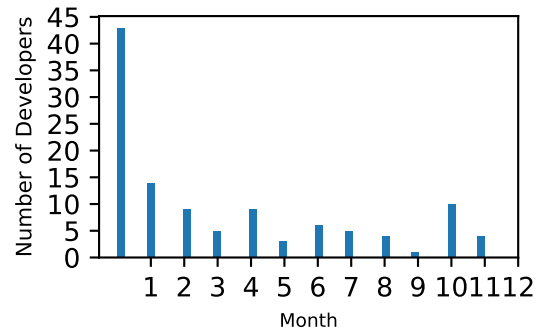
(a) All



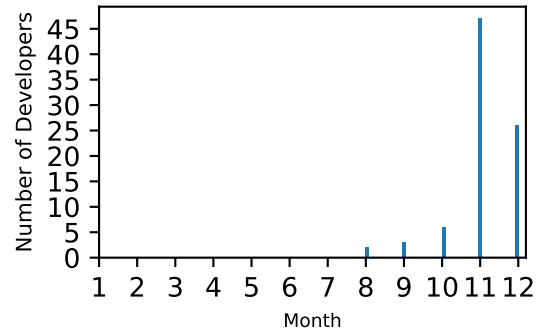
(b) transient founder



(c) sustained founder



(d) transient later joiner



(e) sustained later joiner

Figure 30: A histogram of number of developers that have touched a percentage (90.0%) of methods each month of a year.

3.5 Smooth moving average - 14 For packages touched for each unit

A smooth moving average packages touched on average each period the number of commits to components touched.

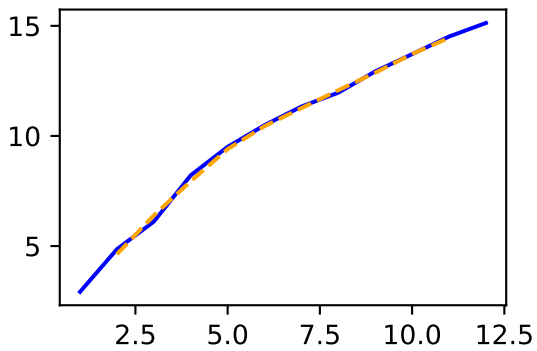
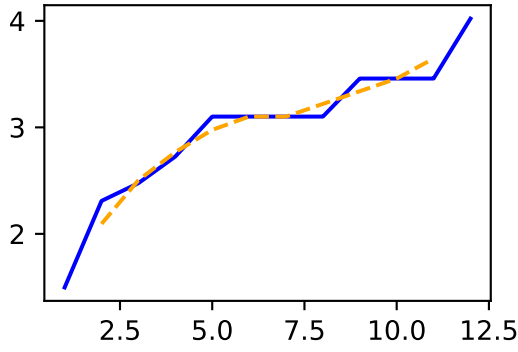
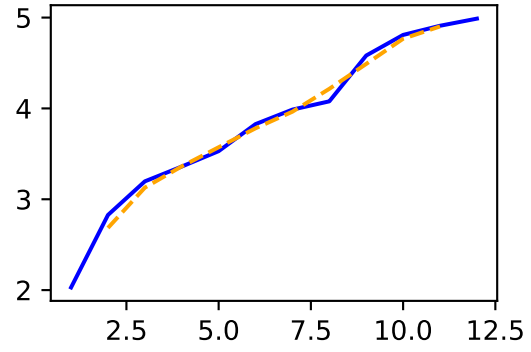


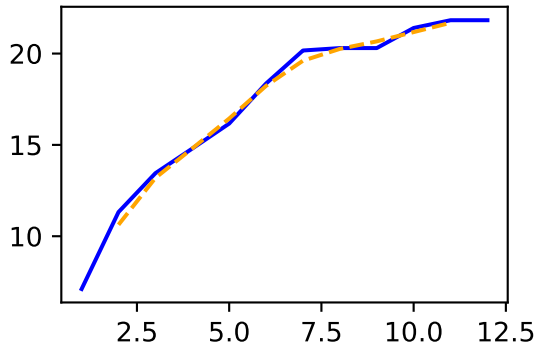
Figure 31: A time series of the number of months (x-axis) against the average (mean) total packages touched (y-axis) for all developers 154.



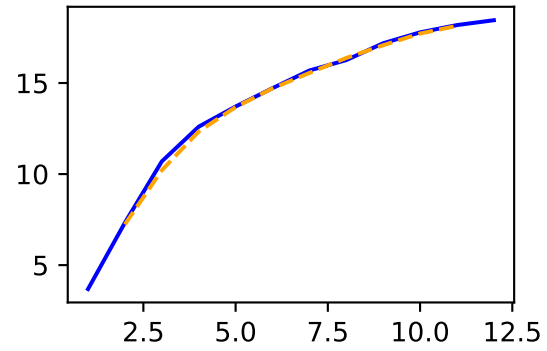
(a) Transient founder developers (8) showing packages touched



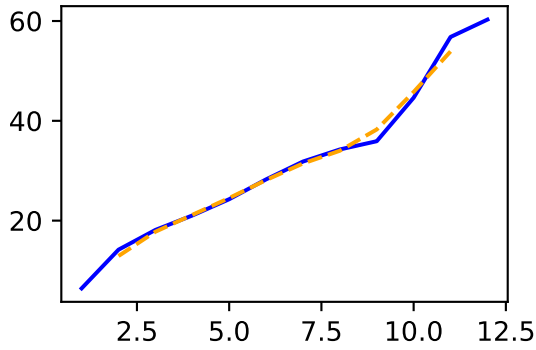
(b) Transient later joiner developers (119) showing packages touched



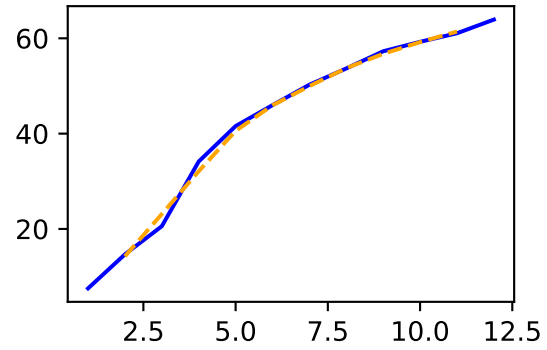
(c) Moderate founder developers (5) showing packages touched



(d) Moderate later joiner developers (85) showing packages touched



(e) Sustained founder developers (6) showing packages touched



(f) Sustained later joiner developers (21) showing packages touched

Figure 32: A smooth moving of total packages touched on average each month of a year.

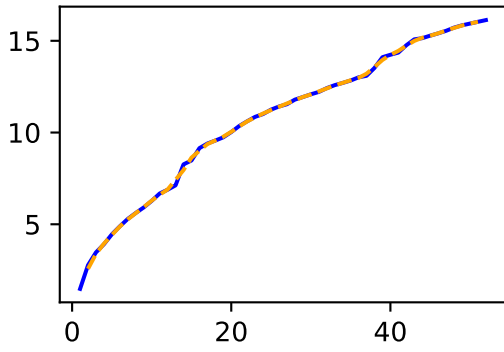
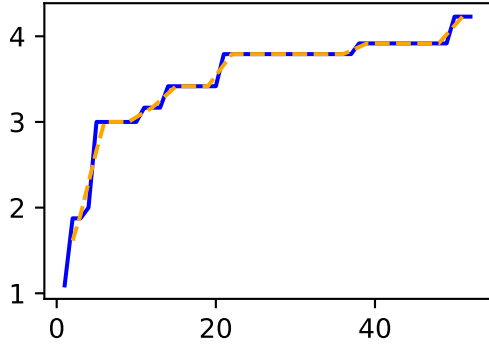
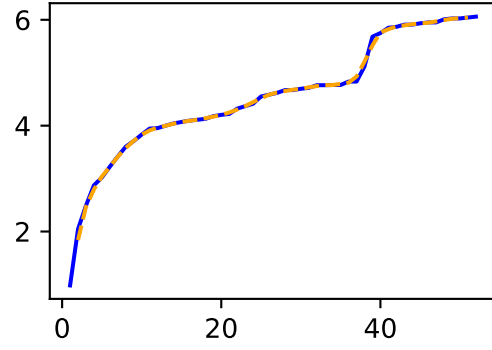


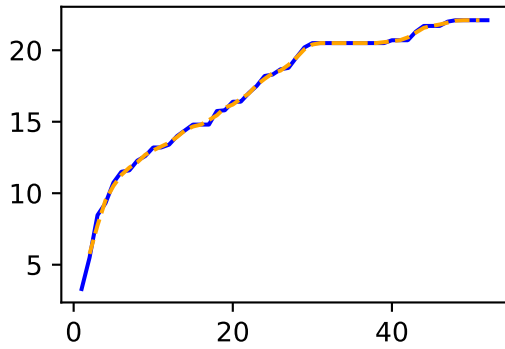
Figure 33: A time series of the number of weeks (x-axis) against the average (mean) total packages touched (y-axis) for all developers 154.



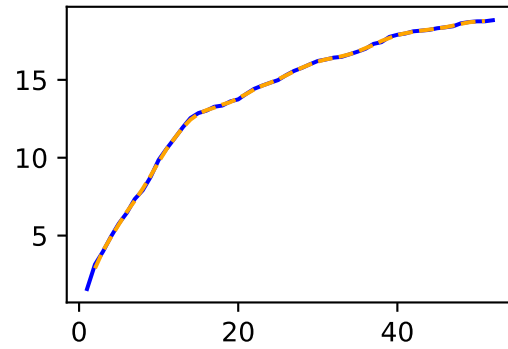
(a) Transient founder developers (8) showing packages touched



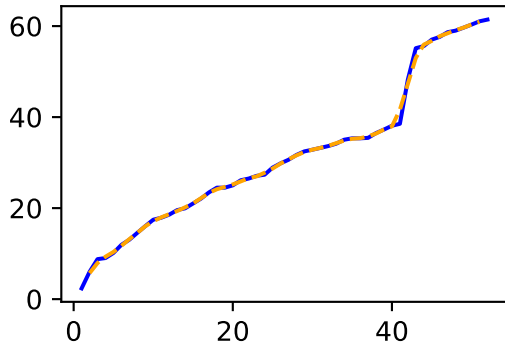
(b) Transient later joiner developers (119) showing packages touched



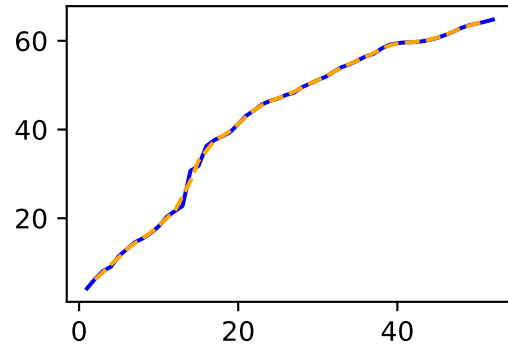
(c) Moderate founder developers (5) showing packages touched



(d) Moderate later joiner developers (85) showing packages touched



(e) Sustained founder developers (6) showing packages touched



(f) Sustained later joiner developers (21) showing packages touched

Figure 34: A smooth moving of total packages touched on average each week of a year.

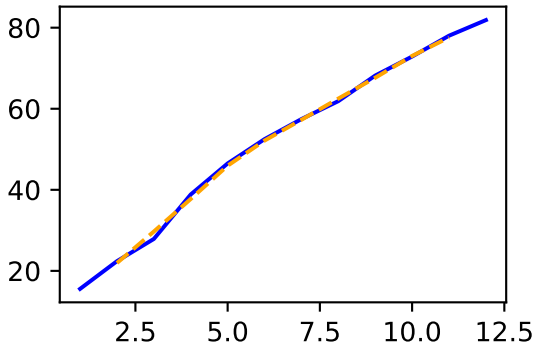
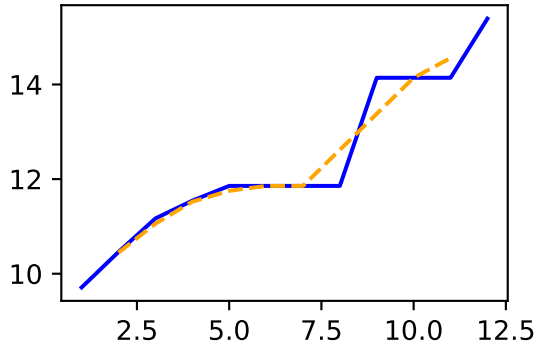
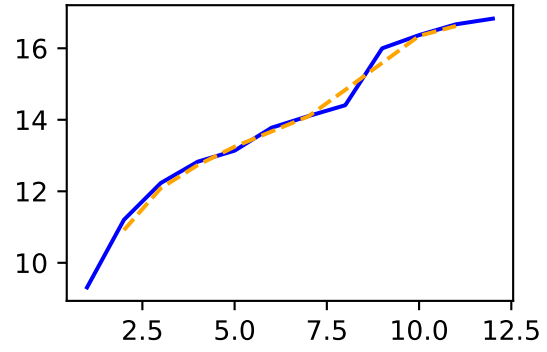


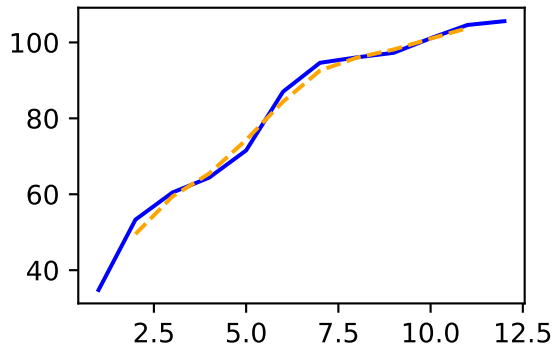
Figure 35: A time series of the number of months (x-axis) against the average (mean) total classes touched (y-axis) for all developers 154.



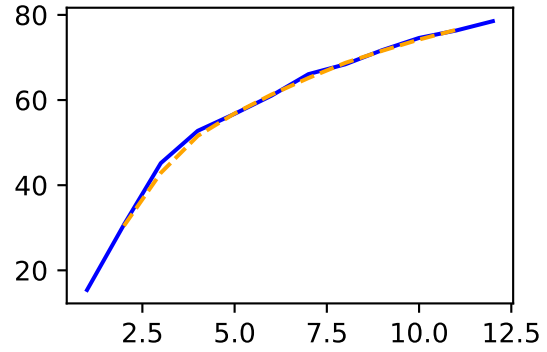
(a) Transient founder developers (8) showing classes touched



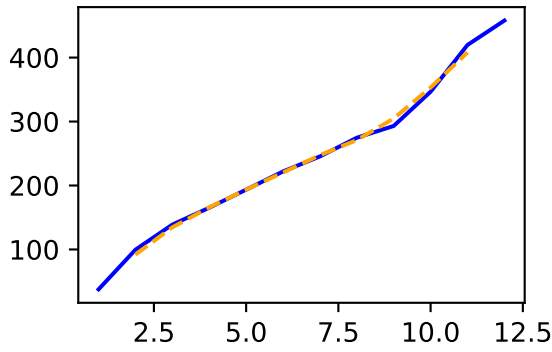
(b) Transient later joiner developers (119) showing classes touched



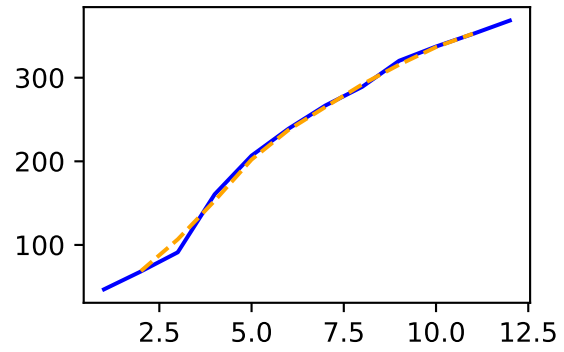
(c) Moderate founder developers (5) showing classes touched



(d) Moderate later joiner developers (85) showing classes touched



(e) Sustained founder developers (6) showing classes touched



(f) Sustained later joiner developers (21) showing classes touched

Figure 36: A smooth moving of total classes touched on average each month of a year.

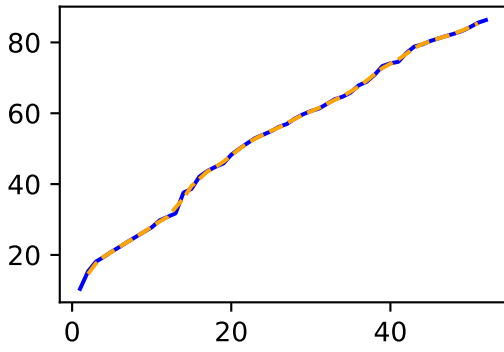
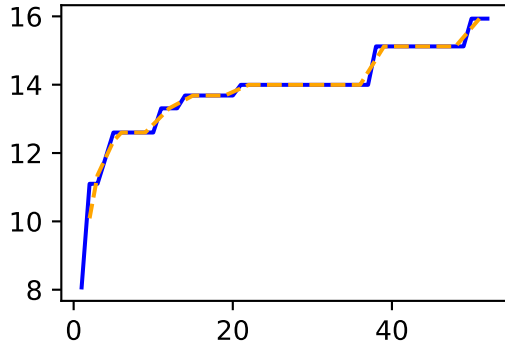
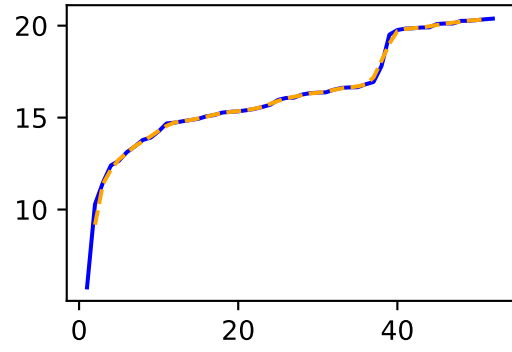


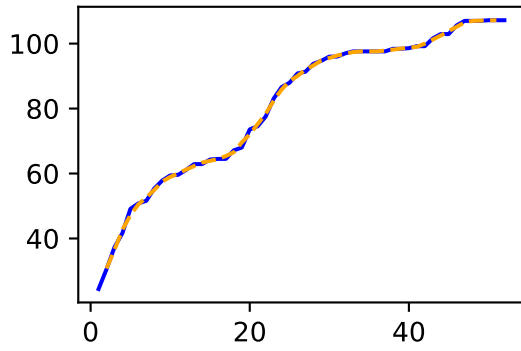
Figure 37: A time series of the number of weeks (x-axis) against the average (mean) total classes touched (y-axis) for all developers 154.



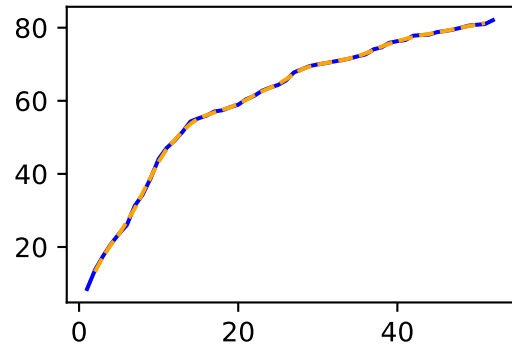
(a) Transient founder developers (8) showing classes touched



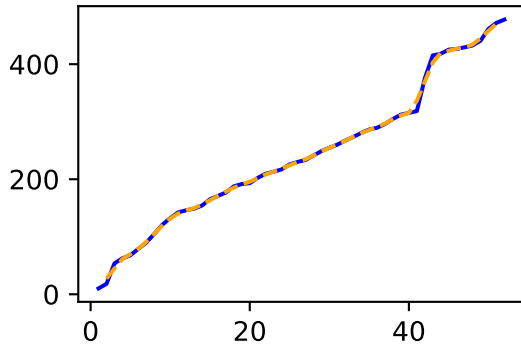
(b) Transient later joiner developers (119) showing classes touched



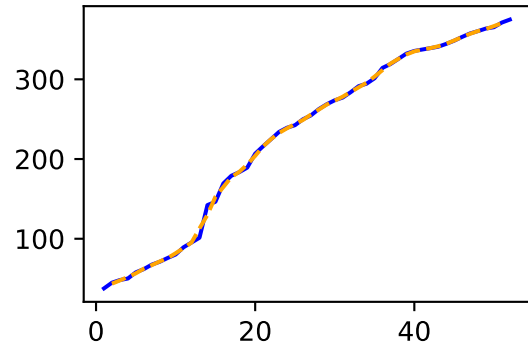
(c) Moderate founder developers (5) showing classes touched



(d) Moderate later joiner developers (85) showing classes touched



(e) Sustained founder developers (6) showing classes touched



(f) Sustained later joiner developers (21) showing classes touched

Figure 38: A smooth moving of total classes touched on average each week of a year.

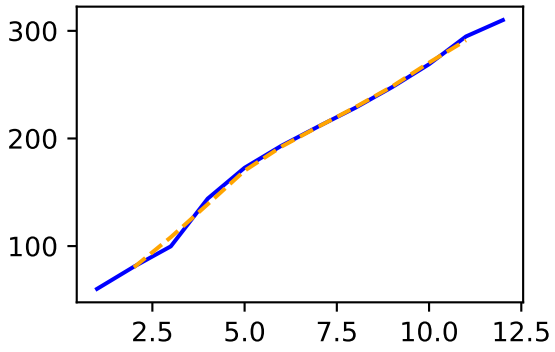
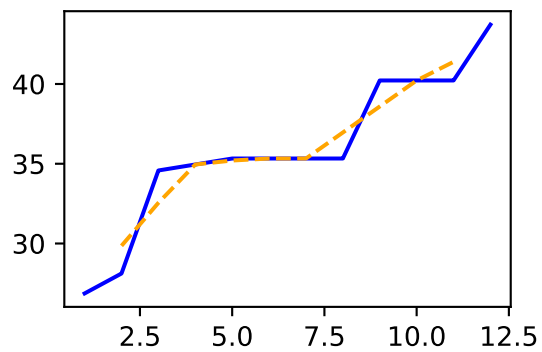
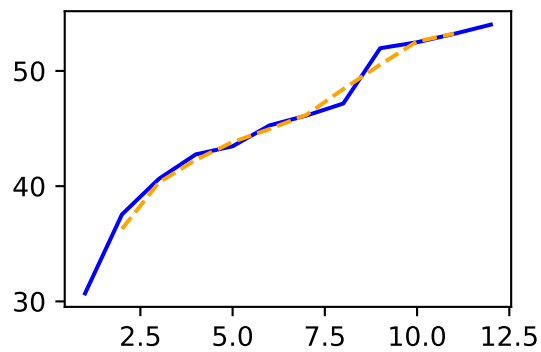


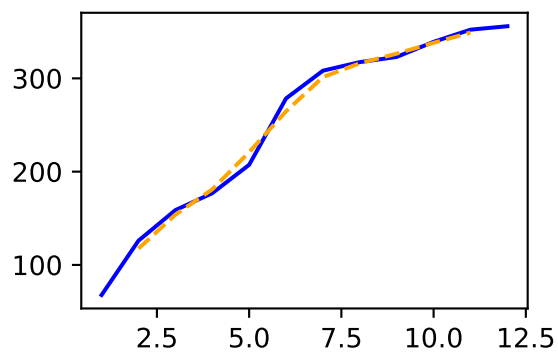
Figure 39: A time series of the number of months (x-axis) against the average (mean) total methods touched (y-axis) for all developers 154.



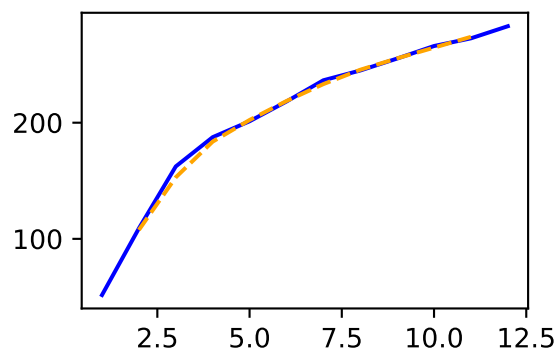
(a) Transient founder developers (8) showing methods touched



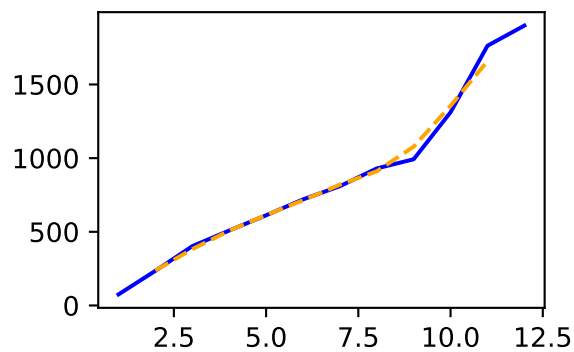
(b) Transient later joiner developers (119) showing methods touched



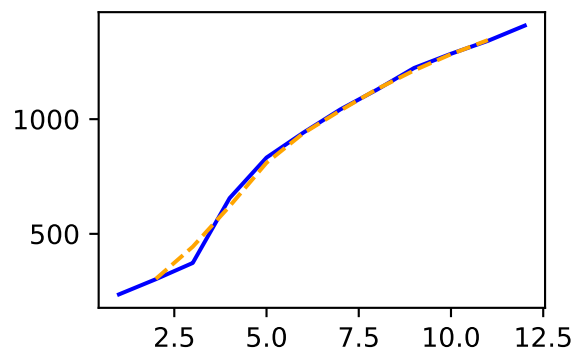
(c) Moderate founder developers (5) showing methods touched



(d) Moderate later joiner developers (85) showing methods touched



(e) Sustained founder developers (6) showing methods touched



(f) Sustained later joiner developers (21) showing methods touched

Figure 40: A smooth moving of total methods touched on average each month of a year.

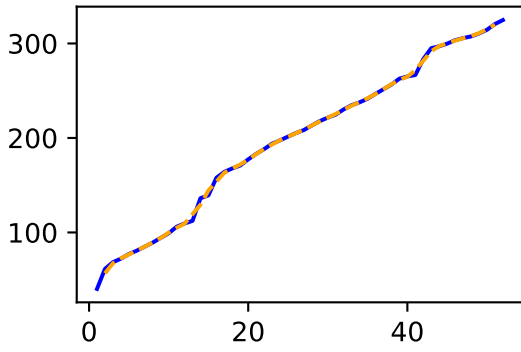
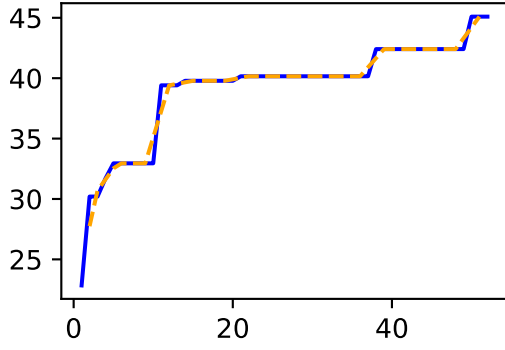
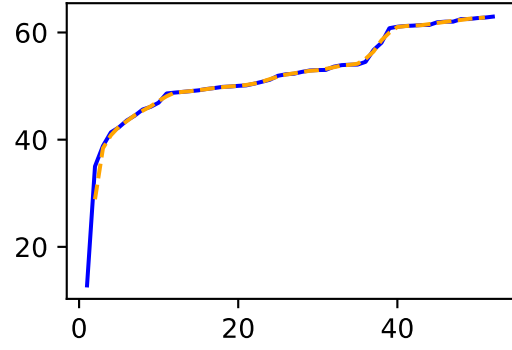


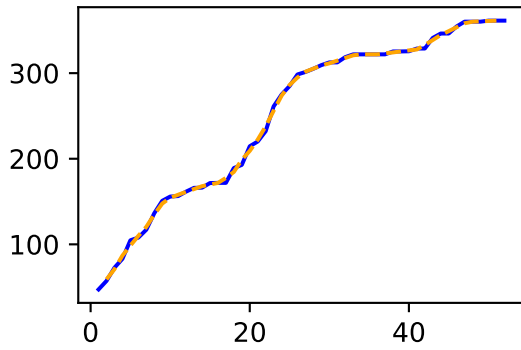
Figure 41: A time series of the number of weeks (x-axis) against the average (mean) total methods touched (y-axis) for all developers 154.



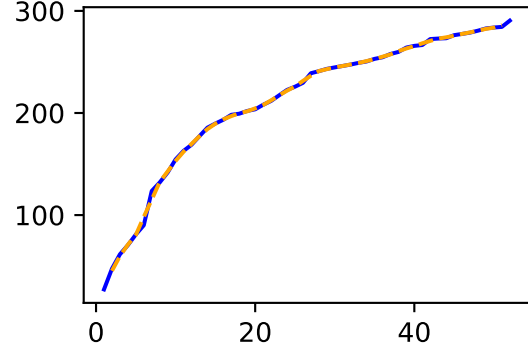
(a) Transient founder developers (8) showing methods touched



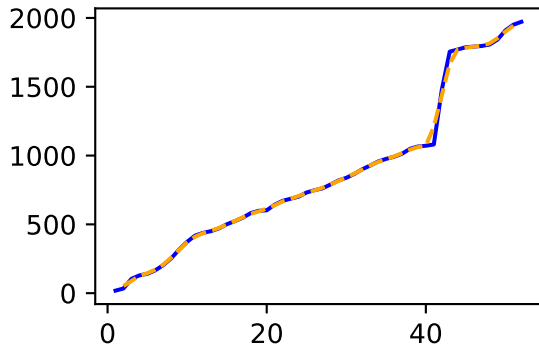
(b) Transient later joiner developers (119) showing methods touched



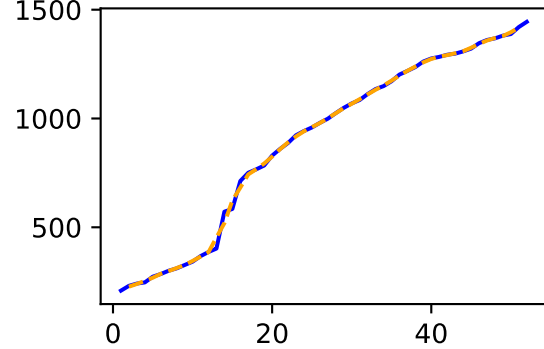
(c) Moderate founder developers (5) showing methods touched



(d) Moderate later joiner developers (85) showing methods touched



(e) Sustained founder developers (6) showing methods touched



(f) Sustained later joiner developers (21) showing methods touched

Figure 42: A smooth moving of total methods touched on average each week of a year.

3.6 Smooth moving average - 21 For packages touched for each unit

A smooth moving average packages touched on average each period the number of commits to components touched.

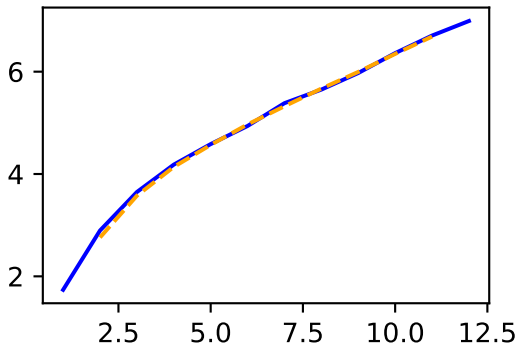
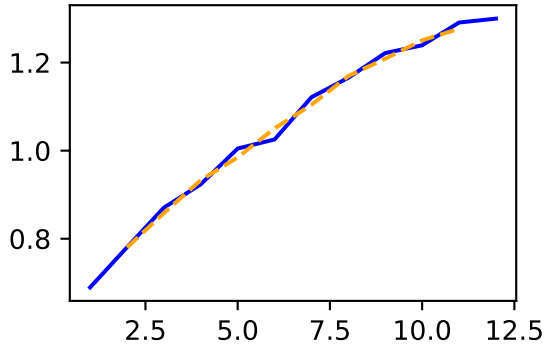
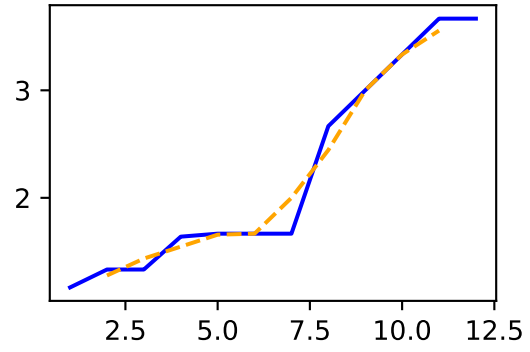


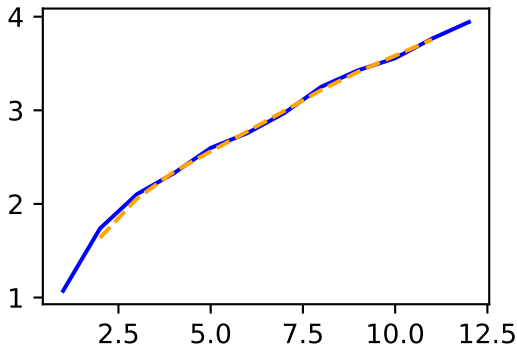
Figure 43: A time series of the number of months (x-axis) against the average (mean) total packages touched (y-axis) for all developers 204.



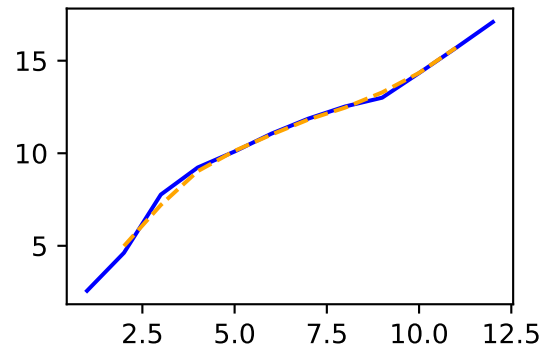
(a) Transient later joiner developers (113) showing packages touched



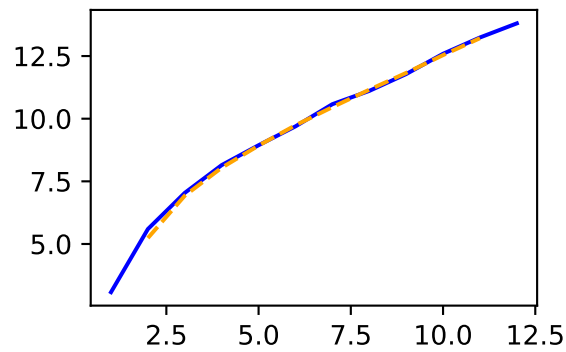
(b) Moderate founder developers (3) showing packages touched



(c) Moderate later joiner developers (93) showing packages touched



(d) Sustained founder developers (7) showing packages touched



(e) Sustained later joiner developers (84) showing packages touched

Figure 44: A smooth moving of total packages touched on average each month of a year.

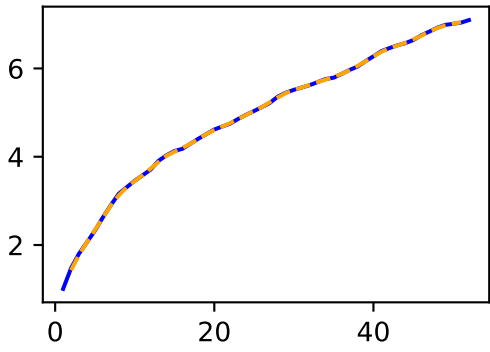
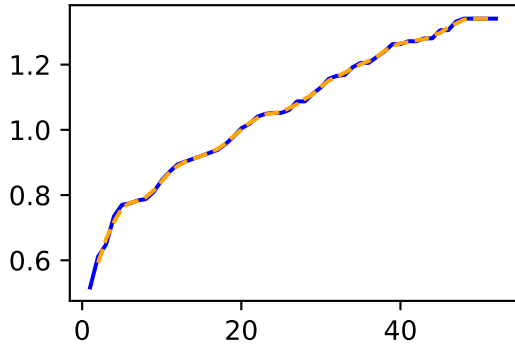
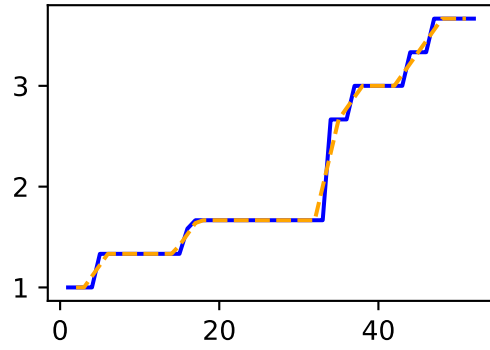


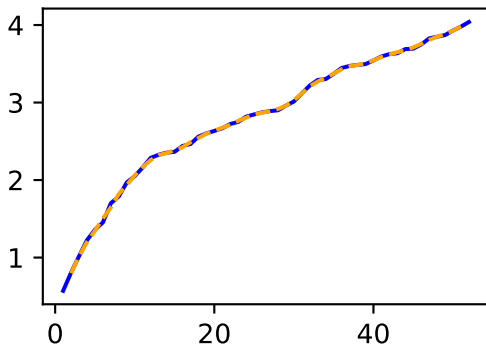
Figure 45: A time series of the number of weeks (x-axis) against the average (mean) total packages touched (y-axis) for all developers 204.



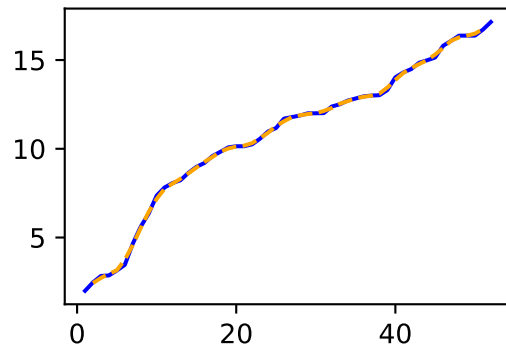
(a) Transient later joiner developers (113) showing packages touched



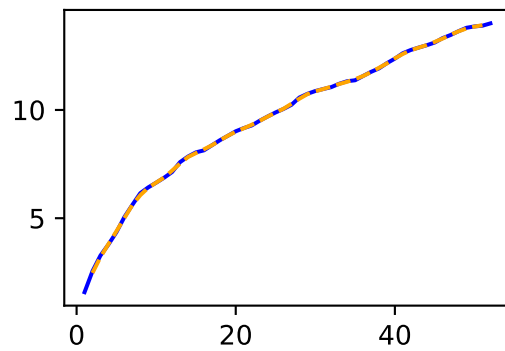
(b) Moderate founder developers (3) showing packages touched



(c) Moderate later joiner developers (93) showing packages touched



(d) Sustained founder developers (7) showing packages touched



(e) Sustained later joiner developers (84) showing packages touched

Figure 46: A smooth moving of total packages touched on average each week of a year.

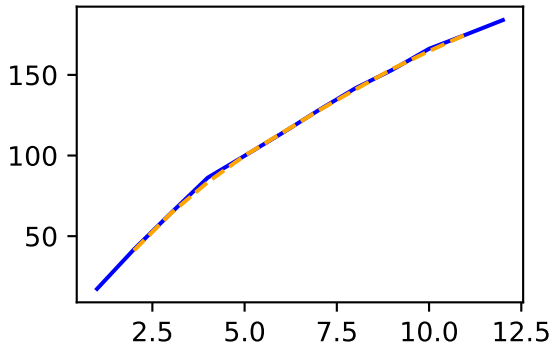
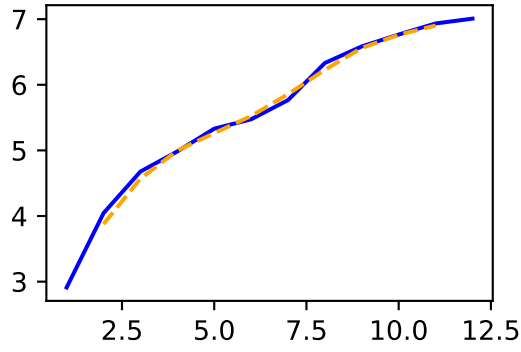
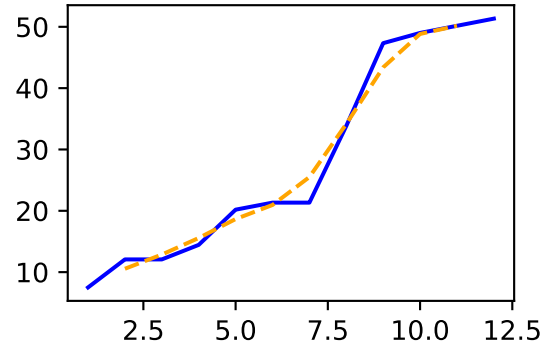


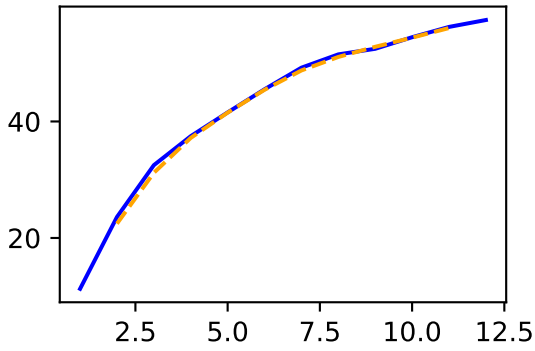
Figure 47: A time series of the number of months (x-axis) against the average (mean) total classes touched (y-axis) for all developers 204.



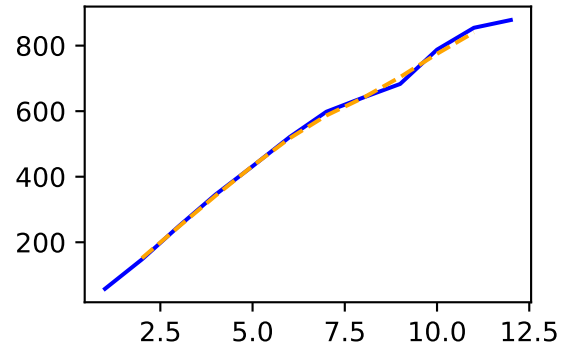
(a) Transient later joiner developers (113) showing classes touched



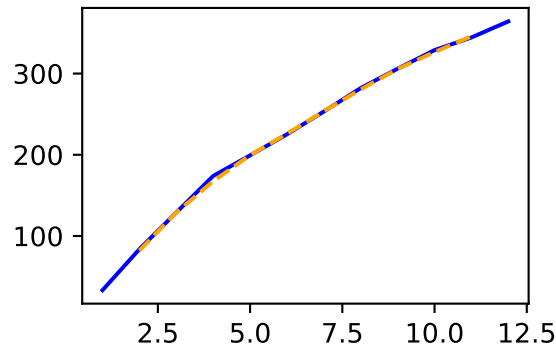
(b) Moderate founder developers (3) showing classes touched



(c) Moderate later joiner developers (93) showing classes touched



(d) Sustained founder developers (7) showing classes touched



(e) Sustained later joiner developers (84) showing classes touched

Figure 48: A smooth moving of total classes touched on average each month of a year.

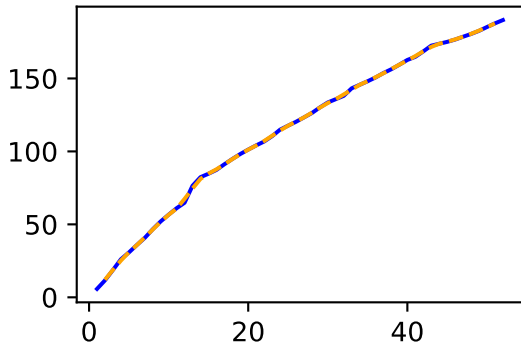
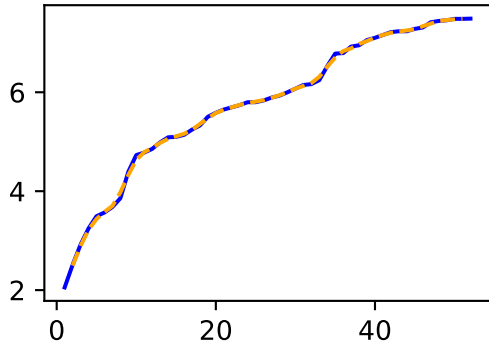
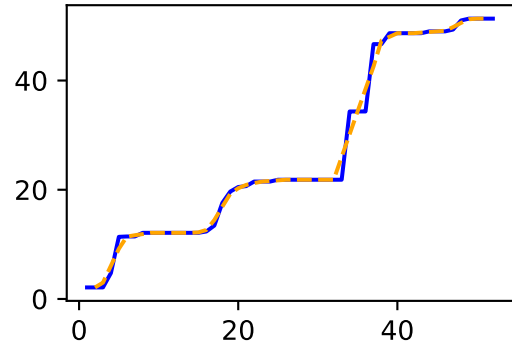


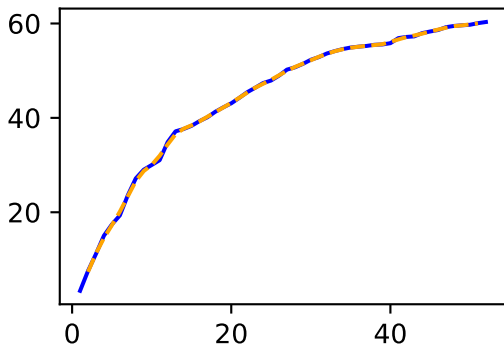
Figure 49: A time series of the number of weeks (x-axis) against the average (mean) total classes touched (y-axis) for all developers 204.



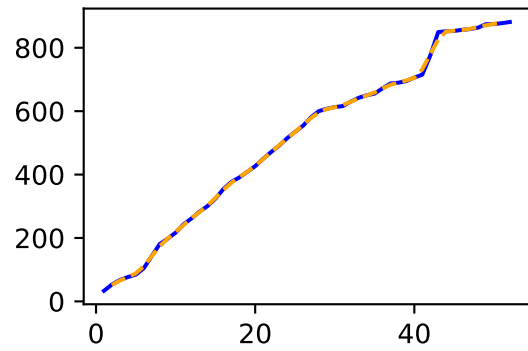
(a) Transient later joiner developers (113) showing classes touched



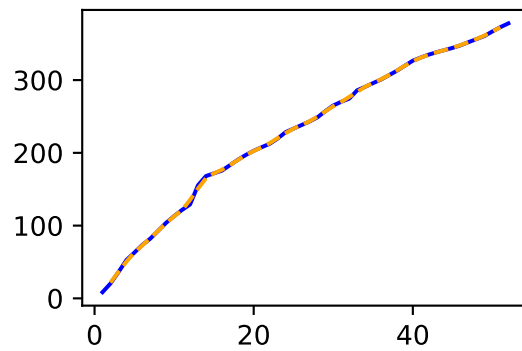
(b) Moderate founder developers (3) showing classes touched



(c) Moderate later joiner developers (93) showing classes touched



(d) Sustained founder developers (7) showing classes touched



(e) Sustained later joiner developers (84) showing classes touched

Figure 50: A smooth moving of total classes touched on average each week of a year.

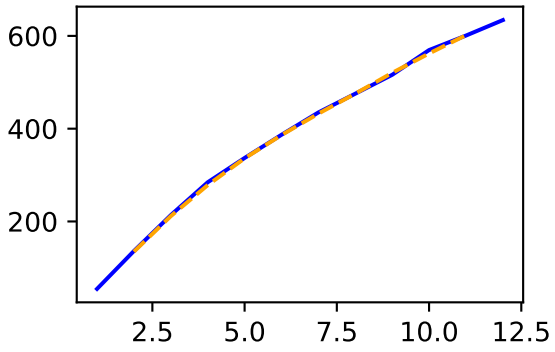
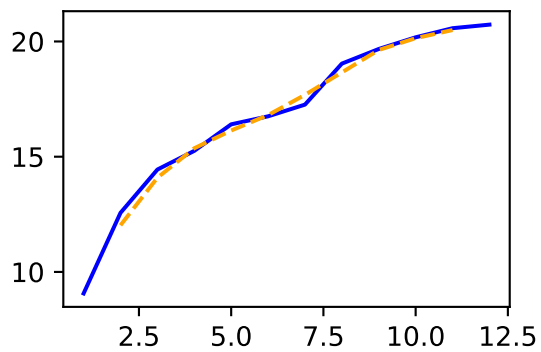
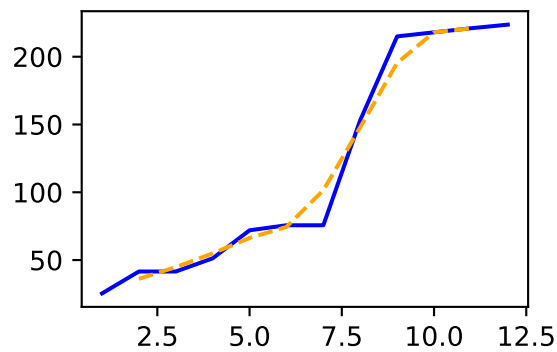


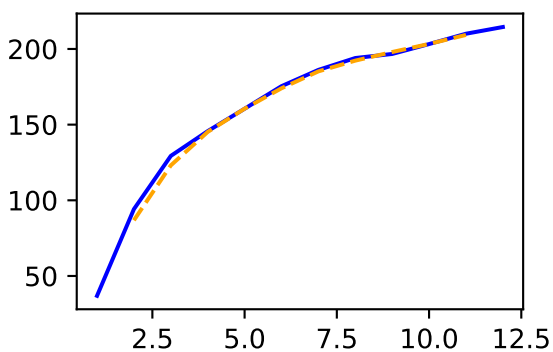
Figure 51: A time series of the number of months (x-axis) against the average (mean) total methods touched (y-axis) for all developers 204.



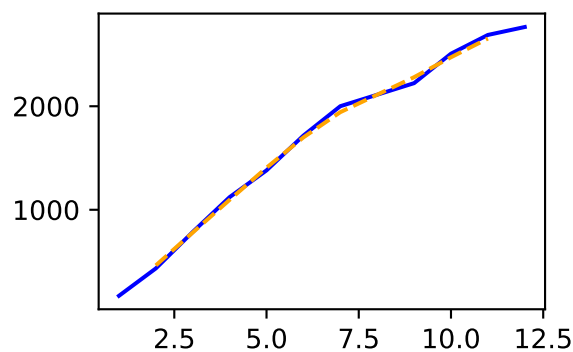
(a) Transient later joiner developers (113) showing methods touched



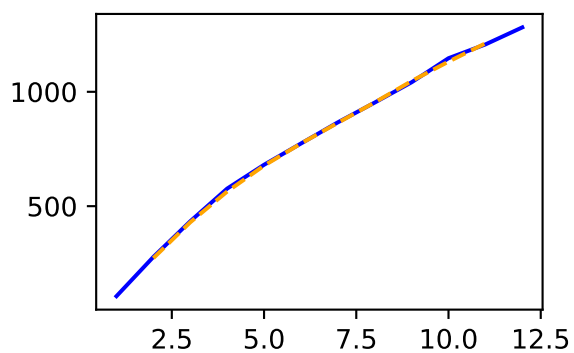
(b) Moderate founder developers (3) showing methods touched



(c) Moderate later joiner developers (93) showing methods touched



(d) Sustained founder developers (7) showing methods touched



(e) Sustained later joiner developers (84) showing methods touched

Figure 52: A smooth moving of total methods touched on average each month of a year.

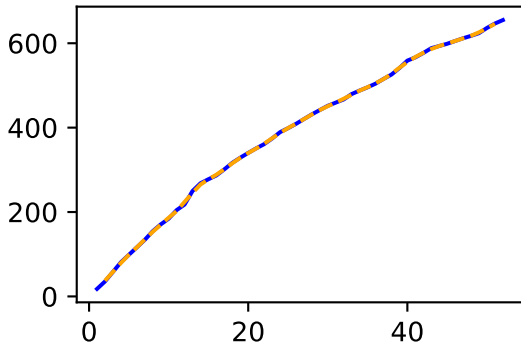
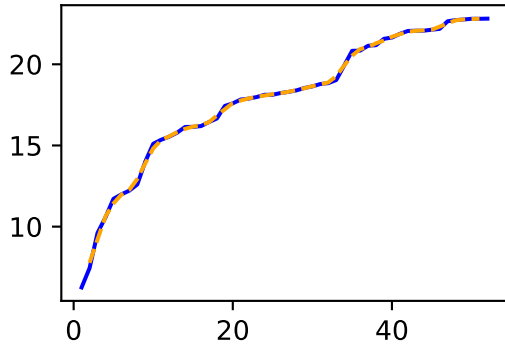
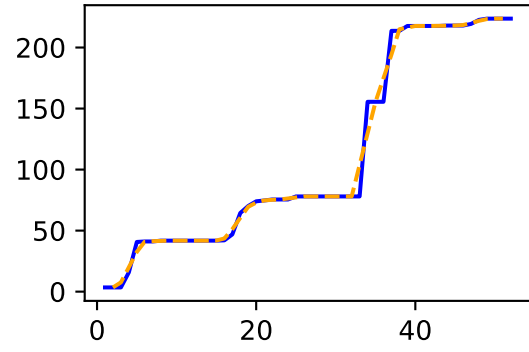


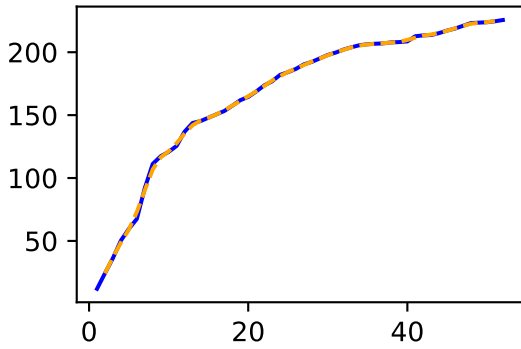
Figure 53: A time series of the number of weeks (x-axis) against the average (mean) total methods touched (y-axis) for all developers 204.



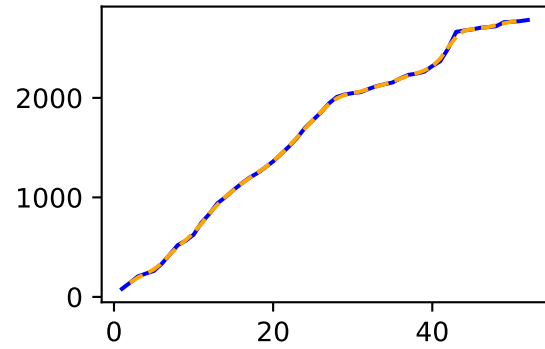
(a) Transient later joiner developers (113) showing methods touched



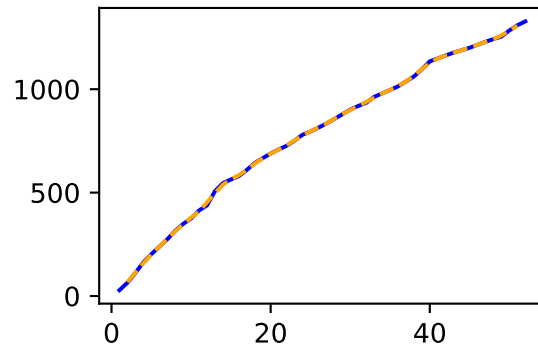
(b) Moderate founder developers (3) showing methods touched



(c) Moderate later joiner developers (93) showing methods touched



(d) Sustained founder developers (7) showing methods touched



(e) Sustained later joiner developers (84) showing methods touched

Figure 54: A smooth moving of total methods touched on average each week of a year.

4 Repository pull request - Pull requests

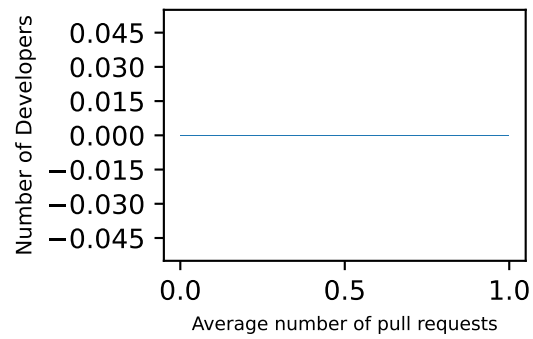


Figure 55: Average number of pull request for 0 developers. The maximum number of developers is 0.